



Information Géographique Volontaire : vers un usage conjoint avec l'information géographique institutionnelle

Ana-Maria Olteanu-Raimond

► To cite this version:

Ana-Maria Olteanu-Raimond. Information Géographique Volontaire : vers un usage conjoint avec l'information géographique institutionnelle. Sciences de l'information et de la communication. Université Gustave Eiffel, 2020. tel-03078869

HAL Id: tel-03078869

<https://hal.science/tel-03078869>

Submitted on 16 Dec 2020

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.

Synthèse des travaux en vue d'obtenir l'Habilitation à Diriger des Recherches délivrée par
l'Université Gustave Eiffel

Spécialité : Sciences et Technologies de l'Information Géographique

Information Géographique Volontaire : vers un usage conjoint avec l'information géographique institutionnelle

Ana-Maria Olteanu-Raimond

LaSTIG, Équipe MEIG

Institut National de l'Information Géographique et Forestière IGN

Soutenue le 23 Octobre 2020

Rapporteurs

Thierry JOLIVEAU, Professeur, Université Jean Monnet de Saint-Etienne

Cyril de RUNZ, Maître de conférences, HDR, Université de Tours

Karine ZEITOUNI, Professeur, Université de Versailles Saint-Quentin

Examineurs

Rodolphe DEVILLERS, DR, Institut de Recherche pour le Développement

Thomas DEVOGELE, Professeur, Université de Tours

Sébastien MUSTIÈRE, Enseignant-Chercheur, HDR, IGN

Anne RUAS, Ingénieur en chef des Ponts, Eaux et Forêts, HDR, IFSTTAR

Remerciements

Une Habilitation à Diriger des Recherches est le fruit d'un travail collaboratif. Ainsi, je souhaite remercier l'ensemble des personnes qui ont contribué à ce travail.

Je souhaite remercier mes trois rapporteurs Thierry Joliveau, Cyril de Runz et Karine Zeitouni d'avoir accepté d'évaluer mon travail. Leurs rapports et leurs commentaires très pertinents et constructifs ont ouvert de longues discussions à la soutenance et de nouvelles perspectives.

Je remercie également les examinateurs Rodolphe Devillers, Thomas Devogele, Sébastien Mustière, et Anne Ruas d'avoir accepté d'examiner mon travail en apportant des commentaires constructifs et bienveillants. Merci à Rodolphe Devillers d'avoir accepté d'être le président du jury et d'avoir égayé la soutenance par visioconférence, en plus de la qualité des remarques et questions.

J'ai soutenu mon HDR 12 ans, à un jour près, après la soutenance de ma thèse. Je tenais à remercier Anne et Sébastien, mes encadrants de thèse, de m'avoir insufflé et transmis leur passion pour la recherche, la rigueur et la persévérance. Merci pour la confiance que vous m'avez accordée il y a 15 ans lorsque vous m'avez accepté en stage, puis en thèse. Nous avons eu l'occasion de retravailler ensemble dans d'autres projets et j'espère que nos collaborations ne s'arrêtent pas ici.

Je remercie les doctorants que j'ai encadrés et ceux que j'encadre actuellement. Merci à Stefan Ivanovic, Mattia Bunel et Ibrahim Maidanah Abdi. Merci pour votre travail, pour les discussions scientifiques fructueuses, et même pour votre ténacité, qui nous fait progresser. Mattia et Ibrahim, j'ai peu discuté de vos travaux dans ce manuscrit ; ceci pour vous laisser l'honneur de défendre votre thèse devant un jury. Bon courage pour vos soutenances de thèse.

Je remercie Cécile, Thomas et Sébastien pour la confiance accordée pour encadrer les thèses de Mattia et Stefan. J'ai beaucoup appris à vos côtés et j'ai eu beaucoup de plaisir à travailler avec vous. Je sais que d'autres collaborations nous attendent. Merci à Arnaud d'avoir accepté d'encadrer la thèse d'Ibrahim.

Merci à Lanfa pour le travail réalisé dans le cadre de son post-doc. Lanfa, tu as accepté notre défi de te lancer dans la théorie des fonctions de croyance sans hésiter. Merci à Laurence et Arnaud pour les discussions fructueuses que nous avons eues ensemble à ce sujet.

Je remercie Marie-Dominique pour les échanges riches que nous avons eus à la fois sur le plan méthodologique et technique. Un grand merci pour ton aide et ton soutien pour préparer la soutenance de mon HDR. Grace à toi, j'ai pu sauver la transmission sur le canal youtube, ce qui a permis de rendre la soutenance publique.

Merci également à Stéphane Bouton pour son aide sur le plan administratif et technique pour préparer la soutenance d'HDR.

Merci à mes collègues avec qui j'ai eu la chance de collaborer dans le projet CHOUCAS : Cécile, Catherine, Mattia, Mickael, Marie-Dominique, Laurence, Sébastien, Véronique.

Dans le cadre du projet LandSense, j'ai eu l'opportunité de travailler et échanger avec plusieurs personnes de l'IGN : Marie-Dominique, Laurence, Arnaud, Ludovic, Julie, Marie, Stéphane, Timothée, Frank, Bénédicte, Pierre, Anna, Cécile, Thierry, Simon, Sylvie... Merci !

Un grand merci à Arnaud, Cécile et Olivier pour la relecture d'une partie ou de la totalité de ce mémoire.

Merci à mes collègues Yann, Sidonie, Guillaume, Bénédicte, Emmanuel pour nos échanges et pour votre soutien ainsi qu'à toutes les personnes qui m'ont témoigné leur sympathie avant ou après la soutenance.

Merci à ma famille de Roumanie qui a assisté à l'intégralité de ma soutenance d'HDR, malgré la barrière de la langue.

Enfin, je tiens à remercier tout particulièrement Olivier, Louise et Gabriel. J'ai rédigé ce mémoire pendant le confinement du printemps 2020 et je vous remercie pour votre compréhension et votre soutien.

Table des matières

Chapitre 1. Introduction	7
1.1 Contexte	7
1.2 Enjeux et objectifs scientifiques	9
1.3 Mon parcours recherche	11
1.4 Démarche globale et organisation du mémoire	14
Chapitre 2. Information géographique volontaire : des pratiques courantes aux recommandations.....	18
2.1 Problématiques et positionnement	19
2.2 Pratiques courantes de l'information géographique volontaire	22
2.3 Recommandations et mise en œuvre d'une plateforme collaborative pour la collecte de l'information géographique volontaire	30
Chapitre 3. Qualification de l'information géographique volontaire.....	40
3.1 Problématiques et positionnement	41
3.2 Proposition d'une méthode de qualification de l'information géographique volontaire issue de plateformes d'autorité.....	44
3.3 Proposition d'une méthodologie de qualification de l'information géographique volontaire issue de plateformes d'activités sportives	57
Chapitre 4. Intégration de l'information géographique hétérogène	65
4.1 Problématiques et positionnement	66
4.2 Identification des mises à jour d'un référentiel du réseau routier d'autorité en utilisant les traces issues des activités sportives.....	69
4.3 Mise à jour de données d'occupation du sol d'autorité	73
4.4 Constitution d'un référentiel métier pour la localisation des victimes en montagne	79
Chapitre 5. Conclusion et perspectives de recherche.....	87
5.1 Information géographique volontaire : des pratiques courantes aux recommandations	88
5.2 Qualification et amélioration de l'information géographique volontaire	92
5.3 Intégrer différentes sources d'information géographique volontaire et institutionnelle en prenant en compte leur hétérogénéité.....	96
5.4 Directions de recherche à long terme	98
Références.....	102
Annexe 1 : Liste de publications.....	113
Annexe 2 : CV.....	117
Annexe 3 : Description des projets.....	121

Chapitre 1. Introduction

Ce mémoire présente une synthèse des recherches que j'ai menées au Laboratoire LaSTIG de l'IGN dans le domaine de la géomatique sur la période 2013 – 2020. Dans ce mémoire, je me concentre sur trois aspects que j'ai traités autour de l'information géographique volontaire : les processus collaboratifs, la qualification et l'intégration de données.

Ce premier chapitre décrit le contexte, les enjeux, les objectifs scientifiques et la démarche générale suivie dans mes travaux de recherche.

1.1 Contexte

La production d'information géographique au niveau national ou européen est traditionnellement réalisée par des institutions publiques ou privées. Ainsi, dans la plupart des pays développés, les Agences Nationales de Cartographie (ANC), telles que l'IGN pour la France, exercent une autorité dans le domaine de la production de l'information géographique et ont comme mission principale la production, la maintenance, la cartographie et la diffusion de référentiels décrivant le territoire et parfois son occupation du sol. Ces référentiels, même s'ils sont différents d'un pays à un autre du point de vue du modèle, de la nomenclature et de la qualité, suivent des processus rigoureux qui assurent l'homogénéité, l'exhaustivité au niveau national, ainsi que la qualité géométrique et thématique.

Au niveau européen, plusieurs initiatives ont vu le jour depuis des décennies telles que Corine Land Cover et Urban Atlas (Büttner, 2014 ; EEA, 2018). La directive INSPIRE, portée par l'Europe, demande l'homogénéisation des référentiels et la diffusion de l'information géographique publique en appui des politiques. L'objectif est de faciliter l'accès aux données institutionnelles à des fins de contrôle, de gestion des risques et de protection de l'environnement et de la bio-diversité.

D'autre part, des entreprises privées s'emparent également de la production de l'information géographique et de la cartographie des territoires dans un but commercial. Nous pouvons citer par exemple Michelin en France, ESRI et Navtech qui ont une portée internationale sans oublier les géants Google et TomTom qui se sont imposés depuis les années 2000. Il convient de préciser que certaines agences de cartographie ont un but commercial, néanmoins ce n'est pas leur principale mission. Mais ce n'est pas l'usage commercial l'information géographique qui nous intéresse dans ce mémoire, mais bien la manière dont **celle-ci est produite, maintenue et utilisée**. Par ailleurs, comme le souligne à juste titre T. Joliveau (Joliveau, 2011), il ne faut pas réduire la production de l'information géographique aux seules institutions publiques nationales et aux entreprises. En effet, de nombreux acteurs publics au niveau régional ou local, des associations non gouvernementales à but non lucratif ont comme mission de produire de l'information géographique, que l'on peut considérer comme institutionnelle. Celle-ci peut être de nature environnementale, statistique, socio-économique, peut décrire le transport de personnes ou de marchandises, etc. Dans le présent mémoire, je m'intéresse à **l'information géographique qui décrit (1) le territoire d'un point de vue topographique et (2) son occupation du sol**.

Les pratiques de production de l'information géographique ont considérablement évolué au cours des dernières décennies. Les progrès technologiques ont eu un impact considérable sur la manière dont l'information géographique est collectée, stockée, analysée et diffusée. En outre, le développement de dispositifs de localisation peu coûteux et l'essor des initiatives du paradigme Web 2.0 ont entraîné une croissance considérable de l'information géographique et ont changé la perception des citoyens par rapport à son acquisition et à son usage.

Ainsi, le citoyen peut d'une manière volontaire ou involontaire produire de l'information géographique. Nous retrouvons dans la littérature le terme de *produser*, qui résulte de la contraction des termes « producer » et « user » et qui désigne le fait que tout citoyen peut être à la fois producteur et utilisateur de l'information géographique (Bruns, 2008). Ce type d'activité a été catalogué premièrement par Turner en 2006 sous le nom de « Neogeography ». D'après Turner, des individus néophytes en géomatique pourraient renouveler les pratiques de production de l'information géographique. Un an plus tard, Michael Goodchild introduit le concept de Volunteered Geographic Information (VGI) qui a pris toute sa place dans la communauté géomatique nationale et internationale et qui est même parfois utilisé à tort, c'est-à-dire que l'usage qui lui est donné ne correspond pas à la définition donnée par (Goodchild, 2007). D'autres termes sont apparus tels que l'information géographique produite par la foule, *crowdsourcing* (Estellés-Arolas et González-Ladrón-de-Guevara, 2012), *user generated content* (Krumm et Narayanaswami, 2008) pour l'information géographique et non-géographique, *user generated spatial content* (Antoniou et al., 2009; Brando et al., 2011) pour l'information géographique.

Associé à une forte volonté d'ouverture de l'information géographique, ce changement de perception et d'usages a révolutionné certains aspects de la cartographie, notamment dans le développement de projets collaboratifs tels qu'OpenStreetMap (OpenStreetMap, 2014) et Wikimapia (Wikimapia, 2014) qui proposent des plateformes et des outils pour l'édition et le téléchargement de données. OSM est de loin le projet collaboratif le plus connu. Sa force réside dans la souplesse du modèle et la motivation des contributeurs pour éditer et améliorer la qualité de l'information géographique. A cet aspect, je peux ajouter que sa notoriété est également due à la communauté de chercheurs qui ont fait de l'écosystème défini autour d'OSM un objet de recherche pour étudier les aspects sociaux, la motivation et les pratiques des contributeurs (Bégin et al., 2018 ; Truong et al., 2019 ; Jokar Arsanjani et al., 2015) sans oublier les nombreuses analyses de qualité et la comparaison avec les référentiels institutionnels (Girres et Touya, 2010 ; Haklay, 2010 ; Ballatore et Zipf, 2015 ; Senaratne et al., 2017). Après une simple recherche dans les titres des publications du mot OpenStreetMap, j'ai recensé 2217 publications dans ScienceDirect, 1241 publications dans ACM Digital et 47 900 publications dans Google Scholar, pour n'énumérer que ces moteurs de recherche académiques. Même s'il peut y avoir un recouvrement des deux autres avec Google Scholar, les chiffres montrent l'intérêt que la communauté de recherche a porté à OSM depuis 2006. Certains travaux, comme c'est le cas des travaux effectués par l'Université de Heidelberg, profitent directement à OSM, puisque les résultats en termes de données et d'outils sont reversés dans l'écosystème OSM¹². Cela est positif et montre l'intérêt du partage des résultats et de la gratuité des données. En dehors de la sphère de la recherche, la notoriété d'OSM est due également à l'utilisation de données OSM par des acteurs locaux ou des entreprises privées comme fond cartographique pour différentes applications web et mobile.

D'autres initiatives, peut-être moins connues, dont le but était de partager une activité ou une passion, ont comme effet l'apparition de nombreux sites web et d'applications où les citoyens partagent des traces enregistrées grâce aux capteurs munis d'un GPS, illustrant leurs parcours sportifs ou touristiques. Il s'agit par exemple des plateformes Strava (Strava, 2020), CamptoCamp (CamptoCamp, 2020) ou VisoRando (VisoRando, 2020).

Enfin, une autre catégorie d'information géographique est celle issue des réseaux sociaux tels que Facebook, Twitter, FourSquare, Flickr et est connue dans la littérature sous le nom de *Ambiant Geographic Information* ou *Social Media Geographic Information* (Stefanidis et al., 2013 ; Campagna et al., 2015). Certes, cette information géographique a été collectée via un acte volontaire de la part des contributeurs, mais l'objectif principal n'est pas de répondre à une initiative de collecte d'information géographique, mais d'exprimer un sentiment, un fait ou de se localiser soi-même ou

¹ <https://osmlanduse.org/#12/8.7/49.4/0/>

² [https://wiki.openstreetmap.org/wiki/University_of_Heidelberg_GIScience_\(Geoinformatics\)_Research_Group](https://wiki.openstreetmap.org/wiki/University_of_Heidelberg_GIScience_(Geoinformatics)_Research_Group)

de localiser une information. Ces pratiques engendrent une hétérogénéité spatiale (*e.g.* zone urbaine vs zone rurale ; zone touristique vs non touristique) et peut présenter de nombreux biais liés aux profils d'utilisateur (*e.g.* les jeunes plus actifs sur les réseaux sociaux). Malgré ces aspects, ce type d'information a attiré de nombreux chercheurs en géomatique et leur usage s'est avéré utile pour étudier les lieux touristiques (Kádár et Gede, 2013), la perception des lieux (Hollenstein et Purves, 2010), pour localiser certains événements (Sakaki et al., 2010) ou pour améliorer les données d'occupation du sol (Yan et al., 2019 ; Sitthi et al., 2016).

Avec l'adoption de la politique d'ouverture des données, connue sous le nom de Open Data, de nombreuses bases de données gouvernementales sont désormais accessibles gratuitement au public. Certaines ANC d'Europe telles que les ANC de la Finlande, des Pays-Bas, de l'Allemagne, de la Grande-Bretagne et de la France ont déjà publié en accès libre une partie ou la totalité de leurs référentiels. Dans la plupart des cas, ces données ont été intégrées dans OpenStreetMap (OSM). Cela a permis d'améliorer l'ensemble des données OSM ou plus rarement d'avoir un retour sur les données ou d'identifier des mises à jour à effectuer dans le référentiel canadien (Bégin, 2014).

1.2 Enjeux et objectifs scientifiques

L'information géographique est omniprésente et en libre accès. Elle n'est pas pour autant facile à appréhender et à utiliser en raison de l'hétérogénéité inter- et intra-sources d'information géographique. L'hétérogénéité est due par exemple au fait que les sources d'information représentent des points de vue différents, sont basées sur des modèles de représentation et de stockage différents, ont des nomenclatures, des dates et des qualités différentes. L'hétérogénéité intra-sources concerne davantage l'information géographique produite par la foule. Etant donné l'absence ou le suivi irrégulier des protocoles de collecte lorsqu'ils existent ou l'usage de capteurs différents, le processus d'acquisition n'est plus maîtrisé et la qualité géométrique ou thématique devient variable. De plus, les métadonnées sont peu détaillées, voire inexistantes. En ce qui concerne le contributeur, malgré son honnêteté et sa motivation, il introduit une part de subjectivité en fonction de son niveau d'expertise, de sa culture ou de sa passion mais également un manque de représentativité spatiale dû à ses centres d'intérêts.

Malgré cette hétérogénéité, l'information géographique produite par la foule offre de nombreux avantages tels que : un volume important, une fraîcheur de l'information et une publication en temps quasi-réel, une qualité croissante, une communauté motivée et prête à participer. Par ailleurs, comme le souligne le rapport Houillé publié en 2016 au sujet des sciences participatives, le collaboratif est « un développement qui a toutes les chances de durer », ... « suit une dynamique mondiale » et ... « n'est pas un effet de mode » (Houiller et Merilhou-Goudard, 2016), les premières initiatives collaboratives datant des années 70. Il recommande, entre autres, de « stimuler le développement des sciences participatives en milieu scolaire et tenir compte de ses spécificités » (Houiller et Merilhou-Goudard, 2016).

L'information géographique produite par la foule soulève ainsi des nombreux défis de recherche pour la communauté en géomatique. Dans mes travaux, je me suis focalisée sur quatre grandes questions de recherche.

- Comment créer et maintenir la motivation d'une communauté de contributeurs autour d'un projet collaboratif ?
- Quel est l'écosystème pertinent pour la réussite d'un projet collaboratif au sein d'une agence nationale de cartographie ?
- Quelle est la qualité de l'information géographique volontaire ? Comment qualifier et améliorer sa qualité ?
- Comment intégrer l'information géographique volontaire pour répondre à un besoin ?

D'autres défis existent bien évidemment tels que la cartographie de l'information géographique volontaire (Touya, 2017), la protection de la vie privée et de la responsabilité civile (Scassa, 2013 ; Hartter et al., 2013), l'anonymisation de données (Shen et al., 2016). Ces défis ne sont pas abordés dans mes travaux de recherche.

Contexte et enjeux pour l'IGN

L'Institut National de l'Information Géographique et Forestière (IGN), en tant que producteur du référentiel à grande échelle (RGE) et de l'occupation du sol sur le territoire français (OCS-GE), a comme mission publique de produire des données géographiques de qualité, exhaustives et actualisées. Avec l'essor des applications et le grand volume d'information géographique disponible, les exigences sont mêmes plus importantes que dans le passé, à la fois de la part de l'Etat mais aussi des utilisateurs (professionnels ou grand public).

L'IGN a été l'un des pionniers en Europe à explorer les aspects collaboratifs dans la production de référentiels. Ainsi en 2008, l'IGN a mis en place le système RIPART (Vigino, 2011), qui permet aux partenaires de l'IGN (*e.g.* municipalités, pompiers, collectivités locales) de remonter de l'information concernant des erreurs ou de nouvelles mises à jour à faire dans le référentiel topographique. Il a ensuite été déployé sur le Géoportail pour un même objectif mais pour un usage grand public. Une version mobile a été proposée aux agents de l'IGN en tant que jeu d'été en 2014. Les résultats encourageants de ces applications pionnières ont conduit l'IGN à proposer un portail citoyen et communautaire unique, nommé Espace collaboratif. Il s'agit d'une plateforme web qui fonctionne sous forme de guichets (un guichet représente une thématique) et permet la remontée d'information et la modification d'une copie du référentiel. Les remontées sont validées manuellement puis traitées par les agents de l'IGN tandis que les éditions sont traitées via des processus de réconciliation et si elles sont validées elles sont introduites dans le référentiel. C'est ce que l'on appelle la mise à jour J+1. L'édition des données géographiques est autorisée uniquement aux partenaires de l'IGN.

En 2018, lors de la redéfinition des orientations principales du projet d'établissement de l'IGN, l'un de ses Ministères de tutelle, le Ministère de la transition écologique et solidaire, a défini l'axe « collaboratif » comme un des trois axes prioritaires de l'IGN.

Dans ce contexte, les enjeux de l'IGN qui impliquent directement mes travaux de recherche sont de :

- Continuer à produire, maintenir et diffuser un référentiel à l'échelle nationale en prenant en compte les exigences d'actualité, d'exhaustivité et d'homogénéité.
- Développer l'axe collaboratif en définissant un écosystème durable qui implique les données, les outils de collecte, les méthodes de qualification et d'intégration, les différentes communautés de contributeurs et d'utilisateurs.
- Bâtir et animer une communauté de contributeurs et d'utilisateurs en développant une relation durable et de confiance.
- Pouvoir qualifier et intégrer le plus automatiquement possible les données provenant de plusieurs partenaires de l'IGN ou opérateurs de l'Etat.

Les travaux de recherche que je décrirai dans le présent mémoire permettent, d'une part, de générer des connaissances sur les processus collaboratifs, sur la motivation des contributeurs, les bonnes pratiques et le potentiel de l'information géographique volontaire. D'autre part, ils fournissent des méthodes de qualification et d'intégration de données hétérogènes qui pourront être adaptées pour le passage à l'échelle.

1.3 Mon parcours recherche

J'ai réalisé ma thèse au laboratoire COGIT de l'IGN sous l'encadrement d'Anne Ruas et de Sébastien Mustière de 2005 à 2008 sur l'appariement des données géographiques. Après la thèse, mes travaux de recherche ont porté davantage sur l'analyse des dynamiques spatio-temporelles à travers deux post-doctorats. Le premier, en 2009, a été effectué au sein du laboratoire Image, Ville, Environnement de l'Université Louis Pasteur de Strasbourg et a porté sur l'analyse des évolutions topographiques à différentes échelles spatio-temporelles. Le deuxième, de 2010 à 2012, a été réalisé au sein du laboratoire OrangeLabs du groupe Orange et a porté sur l'analyse des dynamiques spatio-temporelles à partir des données issues des téléphones portables, que je nommerai dans la suite de ce mémoire l'Information Géographique Involontaire (IGI). Ces travaux, qui ne sont pas détaillés dans ce mémoire, ont donné lieu aux publications suivantes :

- Bonnel, P., E. Hombourger, **A.-M. Olteanu-Raimond** and Z. Smoreda (2017) Apports et limites des données passives de la téléphonie mobile pour la construction de matrices origine-destination, *Revue d'économie régionale et urbaine*, n. 4, pp. 647—671.
- Trasarti, R., **A.-M., Olteanu-Raimond**, M. Nanni, T. Couronné, B. Furletti, F. Giannotti, Z. Smoreda and C. Ziemlicki (2015) Discovering urban and country dynamics from mobile phone data with spatial correlation patterns, *Telecommunications Policy*, vol. 39, n. 3--4, pp. 347—362.
- **Olteanu-Raimond, A.-M.** and T. Couronné (2014) Model the human behavior in space and time by mobile phone data, *Theories and Simulations of Complex Social Systems*, vol. 52, Dabbaghian, V., Mago, V. K. Ed., Intelligent Systems Reference Library, Springer.
- Iovan, C., **A.-M. Olteanu-Raimond**, T. Couronné and Z. Smoreda (2013) Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies, *Geographic Information Science at the Heart of Europe*, pp. 247--265, Springer.
- Smoreda, Z., **A.-M. Olteanu-Raimond** and T. Couronné (2013) Spatiotemporal data from mobile phones for personal mobility assessment, *Transport Survey Methods: Best Practice for Decision Making*, chap. 41, pp. 745--767, Transport Survey Methods: Best Practice for Decision Making.
- Andrienko, G., N. Andrienko, G. Fuchs, **A.-M. Olteanu-Raimond** and J. Symanzik (2013) Extracting Semantics of Individual Places from Movement Data by analyzing Temporal Patterns of visits, 21st ACM SIGSPATIAL International Conference on Advances in GIS.
- Bahoken, F. and **A.-M. Olteanu-Raimond** (2013) Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths; The effect of spatiotemporal filtering on flow measurement, 26th International Cartographic Conference (ICC'13), International conference of ICA.
- **Olteanu-Raimond, A.-M.**, F. Bahoken, T. Couronné and Z. Smoreda (2013) Proposition de matrices de flux temporelles issues de l'activité d'individus mobiles, Actes du Colloque International de Géomatique et d'Analyse Spatiale, SAGEO' 2013.
- **Olteanu-Raimond, A.-M.**, T. Couronné, J. Fen-Chong and Z. Smoreda (2012) Modélisation des trajectoires spatio-temporelles issues des traces numériques de téléphones portables. Le Paris des visiteurs, qu'en disent les téléphones mobiles?, *Revue Internationale de Géomatique*, vol. 22, n. 3, pp. 413--437, doi:doi:10.3166/ri.22.413-437.
- Couronné, T., **Olteanu-Raimond A.-M.** and Z. Smoreda (2011) Looking at spatio-temporal city dynamics through mobile phone lenses, IEEE International Conference of Network of the Future.
- Couronné, T., **Olteanu-Raimond A.-M.** and Z. Smoreda (2011) Urban mobility: velocity and uncertainty in mobile phone data, SocialUrb-2011 Workshop, SocialUrb-2011 Workshop.
- Couronné, T., Z. Smoreda and **Olteanu-Raimond A.-M.** (2011) Chatty mobiles: Individual mobility and communication pattern, Proceeding of Analysis of Mobile Phone Datasets and Networks.
- **Olteanu-Raimond, A.-M.**, R. Trasarti, T. Couronné, F. Giannotti, M. Nanni, Z. Smoreda and C. Ziemlicki (2011) GSM data analysis for tourism application, Proceeding of 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

En janvier 2013, j'ai réintégré le laboratoire COGIT en tant que chargée de recherche de classe 1 du Ministère de l'Ecologie. Dès mon arrivée, j'ai commencé à travailler au sein de deux projets ANR, à savoir Géopeople (voir la description en Annexe 3) et iSpace&Time (voir la description en Annexe 3). Dans le cadre du projet Géopeople, je me suis intéressée aux évolutions des territoires de la fin de 18^{ème} siècle à nos jours, tandis que dans le cadre du projet iSpace&Time, je me suis focalisée sur la

modélisation et la construction de matrices de flux de piétons à partir des traces des téléphones portables. Ces travaux ont donné lieu aux publications suivantes :

- **Olteanu-Raimond, A.-M.** and A. Ruas (2015) Carte de Kohonen et classification ascendante hiérarchique pour l'analyse de données géohistoriques, *Revue Internationale de Géomatique*, vol. 25, n. 4, pp. 581--599, doi:10.3166/RIG.25.581-599.
- Nabavi-Larijani, A., **A.-M. Olteanu-Raimond**, J. Perret, M. Brédif and C. Ziemlicki (2014) Investigating the mobile phone data to estimate the origin destination flow and analysis; a case study: Paris region, *Proceedings of International Symposium of Transport Simulation 2014*
- Ruas, A., C. Plumejeaud, L. Nahassia, E. Grosso, **A.-M. Olteanu-Raimond**, B. Costes, M.-C. Vouloir and C. Motte (2013) GéoPeuple: The creation and the analysis of topographic and demographic data over 200 years, *Advances in Cartography and GIScience, LNG&C, Selection from 26th International Cartographic Conference (ICC'13)*, Buchroithner M., Prechtel N., Burghardt D. Ed., *Lecture Notes in Geoinformation and Cartography*.

De manière naturelle, j'ai élargi progressivement le cadre de mes recherches en m'intéressant à tout type d'information géographique et en abordant toutes les étapes de vie d'une information géographique volontaire, de la modélisation des processus collaboratifs jusqu'à leur intégration pour la production et la mise à jour de référentiels d'autorité en passant par la collecte, la validation et la qualification.

Ces aspects ont été explorés en profondeur grâce à des travaux de thèse ou de post-doctorat que j'ai co-encadrés et à mes travaux de recherche que j'ai réalisés en collaboration avec des collègues du LaSTIG ou à travers mon implication dans différents projets de recherche (*Action COST Mapping and Citizen Sensors*, *Action COST ENERGIC*, le projet H2020 LandSense, le projet ANR Choucas). Les publications issues de ces collaborations ou d'encadrements seront listées au fur et à mesure dans les différents chapitres.

Responsabilités d'administration et d'animation de recherche

Pendant une période de deux ans, de 2014 à 2016, j'ai partagé la co-responsabilité de l'axe de recherche « Qualification et interopérabilité des référentiels de données géographiques » de l'équipe COGIT du LaSTIG avec mon collègue Arnaud Le Guilcher. Pendant cette période l'axe était composé de dix chercheurs (7 permanents et 3 doctorants).

A l'échelle internationale, j'ai été responsable du groupe de travail WG3 *Map Production* de l'*Action COST Mapping and the Citizen Sensor* de 2013 à 2016 (voir la description en Annexe 3). Mon rôle a été d'animer le groupe WG3 composé de 49 membres (chercheurs et représentants de plus de 15 Agences Nationales de Cartographie). L'animation consistait à organiser des réunions et des workshops au sein du groupe, à initier, participer et coordonner les travaux de recherche impliquant les membres du groupe.

A la suite de l'*Action COST Mapping et Citizen Sensors*, j'ai participé au montage du projet H2020 LandSense (<https://landsense.eu/>) pour lequel j'assure la coordination scientifique et administrative pour l'IGN depuis 2016. Le projet LandSense est piloté par IIASA³ et implique 18 partenaires. Il vise à mettre en place un observatoire citoyen pour le suivi de l'occupation et de l'usage du sol en collectant et intégrant l'information géographique produite par la foule. L'objectif de l'IGN dans le projet LandSense est d'évaluer le potentiel de l'information géographique volontaire pour améliorer et mettre à jour les données d'occupation du sol à grande échelle (OCS-GE). Le projet LandSense est présenté plus en détail en Annexe 3.

Depuis 2017, je suis coordinatrice du projet ANR CHOUCAS (<http://choucas.ign.fr/>) qui est une collaboration entre le laboratoire LASTIG de l'IGN (porteur), le Laboratoire d'Informatique de

³ International Institute for Applied Systems Analysis

Grenoble (LIG), le Laboratoire de Mathématique et de leurs Applications de l'Université de Pau et des Pays de l'Adour (UPPA), ainsi que le Peloton de Gendarmerie de Haute Montagne (PGHM) de Grenoble. Le projet CHOUCAS répond à un besoin exprimé par le PGHM de Grenoble. L'objectif du projet est de proposer des méthodes et des outils permettant de constituer et d'enrichir des données géographiques issues de sources hétérogènes, de définir des modèles de raisonnement spatial flou et de visualisation de données imparfaites afin de favoriser le processus de décision dans le contexte de la localisation des victimes en montagne. L'enjeu est donc d'améliorer la capacité de réponse à une alerte. Le projet Choucas est présenté plus en détail à Annexe 3.

Enfin, depuis 2016, je suis impliquée dans le groupe WG IV/3 (Spatial Data Analysis, Statistics and Uncertainty Modelling) de la commission IV de l'ISPRS en tant que secrétaire. Mon activité est à la fois administrative (assurer la communication) et scientifique (participer aux décisions au sein du groupe, participer à l'organisation des conférences, workshops, numéros spéciaux de journaux).

Encadrement. En termes d'encadrement de thèses, j'ai encadré la thèse de Stefan Ivanovic soutenue le 19 janvier 2018 et qui a été co-dirigée par Sébastien Mustière (LaSTIG) et Thomas Devogele (Université de Tours). Le sujet portait sur la mise à jour de référentiel géographique par fusion multi-sources à partir de l'information géographique volontaire (traces GNSS). Actuellement, j'assure la direction de thèse (par dérogation) d'Ibrahim Maidaneh Abdi et l'encadrement de la thèse de Mattia Bunel. La thèse d'Ibrahim Maidaneh Abdi a démarré en 2018 et porte sur l'évaluation de la qualité des données géographiques d'OpenStreetMap (OSM) par apprentissage en vue de définir un référentiel topographique pour la République de Djibouti. Elle est encadrée par mon collègue Arnaud Le Guilcher (LaSTIG). Enfin, la thèse de Mattia Bunel, démarrée en 2017, porte sur modélisation et le raisonnement spatial flou pour l'aide à la localisation de victimes en montagne. La thèse dont la soutenance est prévue à la fin de l'année 2020 est dirigée par ma collègue Cécile Duchêne (LaSTIG).

Etant donné que les thèses de Mattia Bunel et Ibrahim Maidaneh Abdi n'ont pas été encore soutenues, il me semble plus respectueux de leur laisser l'honneur de défendre leurs propositions devant un jury de thèse. Ainsi, ce mémoire n'aborde pas d'une manière exhaustive leurs travaux de recherche. Néanmoins, ils figurent sur la Figure 1.

J'ai également co-encadré deux post-doctorats. Premièrement, le post-doctorat d'Anahid Nabavi-Larijani qui s'est déroulé sur la période 2013-2014 et qui a porté sur la modélisation et la construction de flux origine-destination de piétons à partir de traces des téléphones portables. L'encadrement a été partagé avec mes collègues Julien Perret et Matthieu Brédif (LaSTIG). Je n'aborde pas ces travaux dans ce mémoire. Deuxièmement, en collaboration avec mes collègues Laurence Jolivet et Arnaud le Bris (LaSTIG), nous avons encadré le post-doctorat réalisé par Lanfa Liu sur la période 2019-2020 qui a porté sur l'intégration de l'information géographique volontaire pour la mise à jour de données d'autorité d'occupation de sol.

Enfin, j'ai pu encadrer ou co-encadrer une dizaine de stages depuis 2013. La liste complète des stages est disponible en Annexe 2.

Enseignement. Depuis 2013, je suis intervenue principalement dans le cadre des masters co-habilités par l'ENSG et/ou l'Université Gustave Eiffel (ancienne Université Paris-Est). Mes enseignements, d'une durée d'une quarantaine d'heures par an, portent sur l'analyse statistique spatiale, la qualité et l'intégration de données géographiques hétérogènes, la validation, et les systèmes d'information géographique. Depuis 2015, j'ai la responsabilité du module Analyses statistique de l'Information géographique du master M2IGAST (ENSG-Université Paris-Est). Enfin, je fais partie des jurys de soutenance de stage.

Pour plus d'information sur mon parcours recherche, le CV détaillé est fourni en Annexe 1 et la liste des publications est disponible en Annexe 2.

1.4 Démarche globale et organisation du mémoire

On oppose souvent l'information géographique d'autorité à l'information géographique volontaire. Dans ses propos sur la *Néogeography*, Turner (2006) laisse entendre que les contributeurs néophytes pourraient renouveler les savoirs et les pratiques de la production de l'information géographique et de la cartographie. Des travaux de recherche vont même jusqu'à conclure que l'information géographique volontaire a une qualité égale voire meilleure et pourrait remplacer l'information géographique produite par les ANC (Haklay, 2010 ; Zielstra et Zipf, 2010). Cependant, d'autres travaux de recherche montrent la qualité hétérogène de l'information géographique volontaire (Girres et Touya, 2010), qui serait potentiellement utilisée en tant que source complémentaire pour un besoin sociétal, institutionnel, ou commercial (Haklay et al., 2014 ; Karimipour et Azari, 2015 ; Sedano, 2016 ; Hayat, 2019).

Il me semble que la frontière entre l'information géographique d'autorité et l'information géographique volontaire devient progressivement moins nette. L'information géographique volontaire est de plus en plus utilisée au sein des ANC sous différentes formes. Ainsi, je me suis demandé s'il n'était pas plus judicieux de nuancer les propos évoqués ci-dessus et de considérer que l'information géographique produite par la foule peut être utilisée comme une source d'information complémentaire pour répondre à un besoin institutionnel, d'une part, et que l'expertise et les compétences des experts en géomatique et des néophytes peuvent s'entremêler, chaque contributeur pouvant contribuer en fonction de son expertise (tel contributeur est meilleur pour une tâche de validation, tel autre est plus performant dans l'édition ou la classification des objets géographiques). Afin de mieux prendre en compte les orientations stratégiques et les enjeux de l'IGN, les travaux de recherche que j'ai entrepris ces sept dernières années visent à répondre aux quatre objectifs suivants :

- (1) Identifier les bonnes pratiques et définir des recommandations et des protocoles pour la collecte de l'information géographique volontaire.
- (2) Identifier et définir les mécanismes collaboratifs qui d'une part faciliteront la collecte, la motivation des contributeurs à long terme et l'intégration, et qui d'autre part augmenteront la qualité.
- (3) Qualifier et améliorer l'information géographique volontaire pour faciliter son intégration.
- (4) Intégrer différentes sources d'information géographique volontaire et institutionnelle en prenant en compte leur hétérogénéité.

Pour répondre à ces objectifs et valider les approches proposées, je me suis concentrée sur **trois objectifs applicatifs**. Le premier consiste à étudier le potentiel de l'information géographique volontaire pour la mise à jour de données d'autorité. Je me suis intéressée à deux types de données, les données topographiques de la BDTopo de l'IGN et les données de couverture du sol de la base OCS-GE de l'IGN. L'information géographique volontaire provient à la fois du web et des collectes effectuées avec les outils proposés par l'IGN ou formalisés et développés par nos soins. Le deuxième objectif applicatif consiste à étudier le potentiel de l'information géographique pour constituer un référentiel métier pour la localisation des victimes en montagne à partir des données d'autorité et des sources issues du web. Enfin, le troisième objectif applicatif consiste à étudier le potentiel de l'IGV pour créer un référentiel topographique d'autorité pour les pays dont ces derniers ne sont pas encore disponibles. Ce dernier objectif applicatif n'est pas présenté en détail dans ce mémoire, faisant partie de la thèse d'Ibrahim Maidaneh Abdi.

Le premier objectif applicatif permet d'améliorer et de réduire les coûts des processus de production de l'IGN en contribuant ainsi à la mission de service public. Le deuxième objectif applicatif, quant à lui, permettra de répondre à un besoin sociétal, à savoir sauver des vies.

Les approches proposées s'appuient sur :

- la modélisation des connaissances à l'aide d'ontologies,
- des outils d'intégration tels que l'appariement de données,
- la fusion d'informations en utilisant des méthodes basées sur la théorie des fonctions de croyance ou la logique floue,
- des méthodes d'apprentissage supervisé.

La Figure 1 présente ma vision du workflow de l'information géographique volontaire depuis sa production et jusqu'à son utilisation pour répondre à un besoin donné. L'information géographique volontaire peut-être issue des plateformes proposées par les ANC en s'appuyant sur des mécanismes collaboratifs dont la collecte est maitrisée, d'une part, et de différentes initiatives citoyennes, d'autre part. Elle est qualifiée en utilisant à la fois ses caractéristiques intrinsèques et l'information géographique d'autorité, le cas échéant. Une fois qualifiée, elle peut être utilisée pour répondre aux trois objectifs applicatifs annoncés : la mise à jour, la constitution d'un référentiel métier et d'un référentiel topographique d'autorité. Dans la même Figure 1, nous pouvons également voir le positionnement des différents travaux qui seront décrits dans ce mémoire par rapport au workflow. Les couleurs et les chiffres sont associés aux chapitres.

Ainsi, le mémoire est organisé en six chapitres. Le Chapitre 2 décrit les travaux permettant de comprendre et d'améliorer les processus collaboratif pour la collecte de l'information géographique volontaire, de proposer et d'implémenter un protocole de collecte, et de définir une stratégie collaborative adaptée au contexte institutionnel. Le Chapitre 3 décrit les travaux de recherche portant sur l'évaluation de la qualité de l'information géographique volontaire. Il traite à la fois de l'information géographique volontaire collectée en utilisant nos outils et de celle disponible sur le web. Le Chapitre 4, quant à lui, décrit les approches d'intégration de sources de données géographiques hétérogènes dans le but d'améliorer ou de mettre à jour les bases de données géographiques d'autorité ou de constituer un référentiel métier. Enfin, le Chapitre 5 décrit les conclusions de ces travaux de recherche et ouvrira des perspectives. Sur la Figure 1, les travaux de thèse de Mattia Bunel et d'Ibrahim Maidaneh Abdi, représentés par des traits pointillés, seront discutés dans le Chapitre 5.

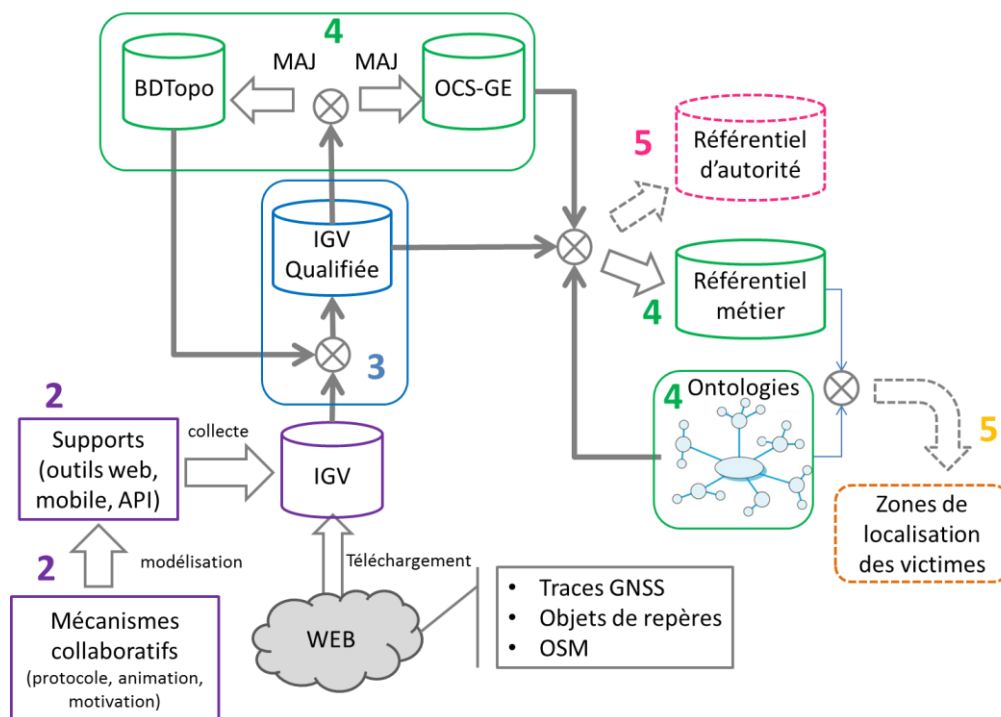


Figure 1. Démarche générale suivie dans mes travaux de recherches

Quelques éléments de vocabulaire

Agence Nationale de Cartographie : institution publique faisant autorité pour la production de l'information géographique.

Alerte guidée : désigne un objet géographique généré soit automatiquement via un algorithme, soit via une intervention humaine, et qui fera l'objet d'une action de la part du contributeur.

Alerte de mise à jour : objet géographique représentant un changement généré via un processus intégrant l'IGV ; il pourrait être intégré dans une nouvelle version de la base de données.

Alerte opportuniste : objet géographique généré par un contributeur et ayant comme objectif de signaler une erreur ou un changement dans une base de données géographiques ou une carte.

Base de données géographiques : représentation sélective et simplifiée du monde réel qui s'appuie sur un modèle et qui suit une méthodologie plus ou moins standardisée pour la production, la maintenance et la diffusion.

Base de données géographiques collaborative : représentation du monde réel qui est issue d'un processus de production impliquant différents contributeurs. La collecte de l'information et sa mise à disposition sont des actes volontaires.

Campagne : évènement impliquant, pendant une durée donnée, la collecte de l'IGV et les activités associées.

Contributeur : individu qui contribue d'une manière active ou passive à la production de l'information géographique.

Donnée géographique : donnée qui est localisée sur la surface du globe.

Information géographique (IG) : désigne l'ensemble des données géographiques qui sont traitées, structurées et organisées ensemble pour répondre à un objectif précis.

Information géographique d'autorité (IGA) : type d'information géographique qui est produite par une autorité tel qu'une ANC en suivant un protocole standardisé.

Information géographique de référence : information géographique qui permet de mesurer ou de contrôler la qualité d'un jeu de données géographiques par comparaison. Elle représente le *terrain nominal*.

Information géographique volontaire (IGV) : type d'information géographique qui est collectée à travers des initiatives collaboratives ayant un but général (*i.e.* décrire le territoire d'un point de vue topographique) ou ciblé (*i.e.* décrire une thématique, un évènement) et qui implique des contributeurs dont l'action est volontaire.

Itinéraire : parcours à suivre entre une origine A et une destination D qui s'appuie sur une infrastructure de réseau physique composée de routes et de sentiers.

Mise à jour : processus qui permet de mettre à jour une base de données géographiques pour actualiser les changements qui ont lieu dans le monde réel.

Mise à niveau : processus qui permet de mettre à jour une base de données géographiques pour corriger les erreurs d'une version antérieure de la base de données géographiques ou pour modifier le modèle de la base de données.

Occupation du sol (OCS) : information géographique décrivant l'occupation du sol. Elle peut être de deux types : la couverture (*i.e.* la vue « physionomique » du terrain décrivant les éléments structurants du paysages tels que le bâti), et l'usage (*i.e.* la vue « anthropique » du territoire qui décrit l'usage des éléments structurants, par exemple le résidentiel).

Objet géographique : objet stocké dans une base de données géographiques qui représente une entité du monde réel et qui est décrit par une géométrie, une sémantique, des attributs thématiques et implicitement par sa forme et sa topologie.

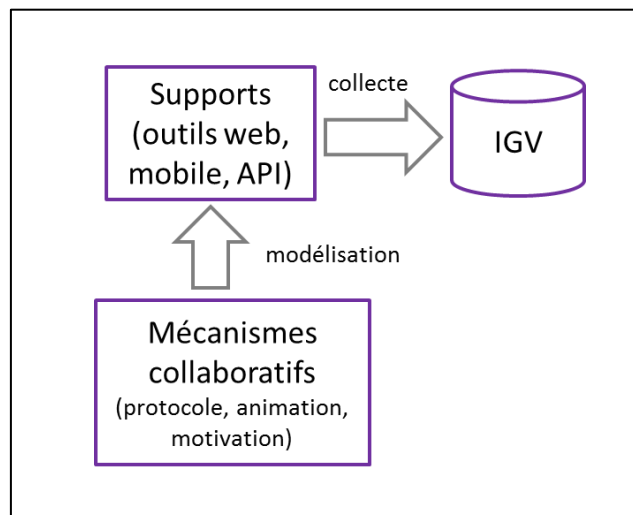
Objet de repère : désigne une entité du monde réel qui peut être utilisée pour se repérer dans un espace. Il peut être représenté par un objet géographique.

Projet collaboratif : initiative pour la collecte de l'information géographique qui permet de répondre à un besoin et qui implique une structure de stockage et de partage d'information, une communauté de contributeurs, et éventuellement des outils de collecte.

Trace GNSS : ensemble de points GNSS (Global Navigation Satellite System) collectés via un capteur basé sur la localisation GNSS et triés selon le temps ; chaque point GNSS est caractérisé par au moins une localisation (longitude, latitude).

Utilisateur : individu qui utilise un outil, un service ou une information géographique volontaire.

Chapitre 2. Information géographique volontaire : des pratiques courantes aux recommandations



Partie de la démarche globale (Figure 1) décrite dans ce chapitre

Ce deuxième chapitre décrit les travaux de recherche qui permettent de répondre aux deux premiers objectifs scientifiques énoncés dans le Chapitre 1 : **(1) identifier les bonnes pratiques et définir des recommandations et des protocoles** pour la collecte de l'IGV et **(2) identifier et définir les mécanismes collaboratifs** qui faciliteront la collecte, la motivation des contributeurs et l'intégration, d'une part, et qui permettront d'augmenter la qualité de l'IGV, d'autre part.

2.1 Problématiques et positionnement

Comme je l'ai déjà évoqué en Introduction, la communauté de recherche en géomatique a accordé un grand intérêt à l'IGV depuis plus d'une décennie. Cette croissance de l'IGV disponible génère plusieurs besoins.

Besoin de comprendre la terminologie

Comme l'a montré (McConchie, 2015), une variété de termes est apparue dans la littérature tout au long de ces dernières années pour décrire l'information produite par la foule. J'ai évoqué en Introduction différents termes qui existent tels que VGI, crowdsourcing, user generated content, iVGI, etc. Malgré leurs différences, ces termes sont souvent utilisés de manière interchangeable pour exprimer l'implication des individus dans l'acquisition de l'information géographique. Les termes sont liés à l'activité collaborative qui est menée d'une manière volontaire ou involontaire et aux objectifs des projets collaboratifs. Par conséquent, en 2013, en partant du constat qu'il n'avait pas de travaux permettant de rassembler ces termes, d'examiner le lien entre eux et d'analyser leur apparition au fil du temps, un premier objectif au sein de l'action *COST Mapping and the Citizen Sensor* a été de compiler les termes rencontrés dans la littérature, de fournir les définitions d'origine et leurs principales attributions et de proposer une classification. Les travaux répondant à cet objectif sont décrits en Section 2.2.1.

Besoin de comprendre les pratiques courantes

La terminologie est liée à l'activité collaborative qui est menée, ainsi qu'aux objectifs des projets collaboratifs. De la même manière que les termes, un grand nombre d'initiatives collaboratives ont vu le jour depuis l'année 2006. Ainsi, un deuxième besoin a été de comprendre les pratiques courantes en termes de types de projet, de motivation, de profils et animation de communautés, de collecte et qualité de l'information géographique.

Concernant **les initiatives des projets collaboratifs**, dans le cadre du projet VGI-net, un premier inventaire avait été publié en 2012 (Elwood et al., 2012) dont l'objectif était de proposer une classification des initiatives allant de la visualisation au partage de données, dans le but final d'étudier leur qualité et leur potentiel pour répondre à des besoins en géomatique. Cependant, cet inventaire était focalisé sur des initiatives locales, certaines n'étant plus d'actualité, et pour la plupart portées par le secteur privé. Ainsi, un besoin d'étendre cet inventaire d'une part, et de rajouter d'autres critères d'analyse d'autre part, a été identifié. Ces travaux sont décrits en Section 2.2.2.

La **motivation des contributeurs** est un élément clé de la réussite d'un projet collaboratif à long terme (Gómez-Barrón et al., 2016). Leurs motivations sont difficiles à identifier avec précision et à généraliser, malgré la littérature existante. Par exemple, Pilz et al. (2006) affirment que la plus grande récompense pour les contributeurs est de voir que les résultats de leurs efforts volontaires de collecte de l'IG sont valorisés. D'autres motivations sont le plaisir et la passion (Chan et al., 2002), ou le sentiment de participer à une cause collective libre telle que l'initiative d'OSM (Jokar Arsanjani et al., 2015) ou celle d'une municipalité française dans les Landes (Larousserie, 2013), ou encore le souhait de répondre à des problématiques telles que les catastrophes naturelles (Goodchild et Glennon, 2010) ou l'accessibilité des personnes à mobilité réduite (Rice et al., 2013), etc. Une autre catégorie de motivation est l'intérêt direct que les contributeurs ont suite à leur contribution. D'une part, les contributeurs, à titre personnel, sont impliqués dans les initiatives publiques car celles-ci peuvent améliorer leur quotidien (e.g. revoir l'implantation des panneaux publicitaires, sources de nuisances visuelles, Sedano, (2016)), ou peuvent leur permettre de bénéficier de subventions (e.g. des contributeurs propriétaires qui ont participé à la cartographie et à la mise à jour du cadastre Basiouka et Potsiou (2012)). Ces initiatives, où les citoyens et différentes communautés sont invités à se prononcer dans les décisions locales, sont également connues sous le nom de *e-Participatory* (Silva, 2013) ou *Public Participation GIS* (Sieber, 2006). D'autre part, l'intérêt peut être économique et implique des agents de compagnies privées. Hayat (2019) expose dans sa thèse ces profils de contributeurs qui participent à l'amélioration des données OSM bien identifiées en amont, et qui par

la suite sont utilisées dans des chaînes de production telles que des cartes touristiques pour Michelin ou des fonds de carte pour MapBox. La motivation peut être aussi de nature sociale. Ainsi, dans les réseaux sociaux (Flick'R, Instagram, Facebook, Tweeter), partager des informations sur les lieux visités, les déplacements, ses opinions ou la compréhension d'une situation commune est devenue une pratique très courante. Pour un état de l'art sur la motivation des contributeurs, le lecteur peut se référer à (Budhathoki et Haythornthwaite, 2013; Fritz et al., 2017).

Un dernier aspect qui peut être source de motivation est l'éducation et l'opportunité pour s'instruire. A titre d'exemple, nous pouvons citer le portail citoyen Vigi-Nature⁴ porté par le Muséum National d'Histoire Naturelle qui présente, entre autres, un axe important lié à l'éducation, et l'Edu-Géo⁵ de l'IGN qui propose des outils pour l'édition et la cartographie collaborative. Récemment, l'ANC de Finlande a développé l'application #hylo dont l'objectif est d'introduire les sciences participatives dans les écoles (Halkosaari et al., 2017). Le même programme a été transposé par Kadaster de Pays-Bas (Grus et Laakso, 2016 ; Bol et al., 2016).

Le profil des contributeurs et les différentes communautés ont fait également l'objet de plusieurs travaux de recherche. Ainsi, nous avons appris que le profil typique d'un contributeur OSM est un homme, éduqué, et ayant des compétences techniques, et que les hommes participent davantage pour des intérêts privés, alors que les femmes contribuent dans le contexte de leur travail (Schmidt et Klettner, 2013). La qualité de l'IGV varie selon l'expertise des contributeurs ou la fréquence de leur contribution. Par exemple, Coleman et al. (2009) classifient les contributeurs selon leur expertise en cinq catégories : néophyte, amateur intéressé, expert amateur, expert professionnel et autorité experte. Truong et al. (2019) proposent, en revanche, une classification basée sur l'interaction des contributeurs en trois catégories : pionniers, modérateurs et véridiques. Heipke (2010) propose d'élargir la classification de Coleman et al. (2009) en se basant sur le type de contributions telles que les activités sportives, les réseaux sociaux, la cartographie ouverte, etc. C'est justement cette classification qui nous intéressera pas la suite, dans un contexte institutionnel.

Concernant la fréquence de participation, il y a une inégalité au sein des communautés en ligne et des réseaux sociaux (Capineri et al., 2016), comme détaillé dans le chapitre 3. Par exemple, la fameuse loi de l'inégalité '90-9-1' signifie que 90% du contenu est généré par 1% des utilisateurs (Nielsen, 2016). Parmi les autres contributeurs, 9 % fournissent du contenu de temps en temps, tandis que 90 % utilisent le contenu mais n'apportent aucune contribution. Pour OSM, la loi est estimée à '81-14-5' (Neis et al., 2013). Une pratique très courante pour impliquer les contributeurs est le concept de *carto-party* (en anglais mapathons) où les contributeurs se donnent rendez-vous pour répondre à un objectif précis. Cette pratique instaurée par OSM notamment via l'association OSMHot (Humanitarian OSM) a été élargie dans d'autres communautés également. La *gamification* est un autre concept, dont l'efficacité a été prouvée, pour créer une communauté de contributeurs. A titre d'exemples, nous pouvons évoquer les travaux de l'équipe de IIASA sur les plateformes Geo-Wiki, Laco-Wiki, Photo-Quest-Go, Picture-Pile qui permettent la collecte de l'IGV pour la constitution de données de validation pour les données de couverture et d'usage du sol (Fritz et al., 2012 ; See et al., 2017; Fritz et al., 2017).

Nous avons vu ainsi que de nombreux travaux de recherche se sont intéressés à étudier la motivation des contributeurs de la communauté OSM ou de la communauté des sciences participatives (e.g. protection de la biodiversité, observation de la flore et de la faune). Nous en savons peu sur la motivation des autres communautés. Naturellement, nous pouvons par exemple émettre l'hypothèse que pour la communauté des pratiques sportives, la motivation est directement liée à l'intérêt, la passion. En revanche, pour la communauté impliquée dans les projets collaboratifs menés par les ANC, il est difficile d'identifier la motivation. De plus, malgré les travaux effectués au

⁴ <http://www.vigienature.fr/>

⁵ <https://www.edugeo.fr/>

sein de la communauté EuroSDR⁶ (Mooney et Morley, 2014), nous en savons peu sur l'usage réel de l'IGV au sein des ANC.

Ainsi, trois questions se posent : Quel est l'usage de l'IGV au sein des ANC ? Quel en sont les freins ? Quelle est la motivation des contributeurs potentiels autour d'une plateforme collaborative proposée par une ANC ? Ces aspects sont traités en Section 2.2.3.

Besoin de protocoles de collecte

Alors que les ANC utilisent des protocoles robustes et standardisés, nommés spécifications, qui guident la collecte de données géographiques, les projets collaboratifs manquent souvent de protocoles ou proposent des recommandations plutôt que des spécifications strictes. Par exemple, la communauté OSM dispose d'un wiki qui s'enrichit progressivement de recommandations concernant l'édition des objets ou l'attribution de clés « key » (Antoniou et Skopeliti, 2017). L'agence nationale United States Geological Survey (USGS), qui a été une des premières ANC à ouvrir les données aux contributeurs volontaires, propose également des guides d'utilisation sur sa page web⁷, mais pas de protocoles. D'autres, en revanche, proposent des protocoles pour assister les contributeurs afin de collecter de l'IGV de très bonne qualité et qui permette de répondre à des objectifs bien définis. Par exemple, dans les sciences participatives, dont les premières initiatives remontent aux années 70, le protocole de collecte est présent dans la quasi-totalité des projets collaboratifs (Bonney et al., 2009; Pocock et al., 2014). Dans le domaine de la géomatique, nous pouvons citer les protocoles définis dans la plateforme Géo-wiki (Fritz et al., 2012, 2009) dont le rôle est de guider le contributeur dans la classification de la couverture des sols ou dans la prise de photos *in-situ*, ou encore les travaux de Geo-Historical Data (Perret et al., 2015) et du projet Géopeople (Ruas et al., 2014) qui proposent des protocoles pour vectoriser des cartes anciennes.

Bien que l'IGV puisse potentiellement atteindre des normes de qualité élevées sans protocoles rigoureux, comme il a été prouvé dans différents travaux (Haklay, 2010 ; Zielstra et Zipf, 2010 ; Neis et al., 2013 ; Girres et Touya, 2010), leur absence est souvent une source d'imperfections et d'hétérogénéités (Follin et al., 2019) et peut représenter un obstacle à leur réutilisation (Johnson et Sieber, 2013). L'absence de protocoles génère des problèmes d'incohérence même pour des contributeurs avérés, comme cela a été le cas lors de l'intégration de Corine Land Cover dans OSM (Haklay et al., 2014; Touya et Brando-Escobar, 2013). Pour les contributeurs néophytes qui participent occasionnellement, cela peut engendrer des biais (Antoniou et Schlieder, 2018) ou des hétérogénéités dans l'usage de toponymes lorsque des dialectes existent dans la région d'étude (Haklay et al., 2014).

La nécessité d'établir des normes et des protocoles pour les projets collaboratifs n'est pas une nouveauté. De nombreux auteurs, dont Sui (2007) et Johnson and Sieber (2013), ont souligné la pertinence des protocoles dans les projets collaboratifs et les problèmes potentiels qui se posent pour les communautés et la société lorsque ceux-ci sont absents. D'autres travaux ont montré que l'existence des protocoles a un impact direct sur la qualité de l'IGV collectée (Bonney et al., 2009; Foody et Boyd, 2012; Pocock et al., 2014).

Conscients de ces pratiques et initiatives actuelles, la question de savoir pourquoi une ANC devrait également proposer une plateforme VGI est pertinente. Le rôle potentiel des citoyens dans les missions des institutions gouvernementales a été identifié (Antoniou et al., 2009 ; Mooney et Morley, 2014 ; Haklay et al., 2014) mais les auteurs n'ont pas mis l'accent sur l'utilisation réelle de l'IGV au sein des ANC.

Hypothèses de recherche. Il convient donc d'accorder une plus grande attention d'une part à l'étape de collecte de données, afin de garantir la qualité de l'IGV collectées, et d'autre part à la création de

⁶ <http://www.eurosd.net/about>

⁷ http://navigator.er.usgs.gov/help/vgistructures_userguide.html

la communauté et de son animation lorsqu'un projet collaboratif doit être mis en place. Notre première hypothèse de recherche est basée sur le fait que, comme différents auteurs l'ont montré, **la qualité de l'IGV peut être améliorée grâce aux protocoles de collecte et que maîtriser la collecte permet de maîtriser la qualité**. Une solution serait donc d'élaborer un protocole de collecte qui définit les grandes lignes directrices, **génériques et flexibles** pour la collecte de l'IGV et indépendamment d'un projet spécifique de collecte d'IGV. Une attention particulière devrait être accordée à la structure et à la complexité de ces protocoles. Premièrement, ils doivent garantir la simplicité afin de s'assurer que les contributeurs ne seront pas découragés par la complexité de la mise en œuvre. Il faut en effet éviter de proposer un protocole qui provoque la frustration ou la démotivation des contributeurs. Deuxièmement, ils doivent garantir que l'IGV collectée présente une qualité acceptable pour les besoins du projet collaboratif donné.

Le protocole générique proposé est décrit en 2.3.1 et son implémentation dans le cadre d'un projet collaboratif ayant comme objectif la mise à jour des données d'occupation du sol est présentée en Section 2.3.2.

Notre deuxième hypothèse de recherche est que **les profils des contributeurs, malgré leur hétérogénéité, peuvent se compléter à condition d'identifier des objectifs précis adaptés aux expertises et aux motivations de chaque profil**. Ainsi, il est important de bien identifier la motivation des contributeurs, qui sont les contributeurs et quels sont les mécanismes qui permettent de créer une communauté durable. Bien que cet aspect n'ait pas été un sujet central dans nos travaux, nous avons utilisé les connaissances issues de l'état de l'art et notamment les travaux de Tweddle et al. (2012) qui proposent différentes stratégies pour créer une communauté, où le point de départ est de déterminer le public cible : grand public, amateurs, experts de l'IG, étudiants, élèves, etc. Cet état de l'art a été complété avec une étude réalisée avec ma collègue Laurence Jolivet pour mieux connaître la communauté des professionnels, partenaires de l'IGN (voir la Section 3.2.1.2).

2.2 Pratiques courantes de l'information géographique volontaire

Avant-propos

Cette section présente les travaux que j'ai réalisés en collaboration avec des collègues chercheurs impliqués dans l'Action COST *Mapping and Citizen Sensors*. Les partenaires privilégiés ont été : IIASA, l'Université de Coimbra, de Milano, d'Athènes, Nottingham et de Maynooth, ainsi que les ANC de Finlande, Suisse, Pays-Bas, UK. Un premier aspect consiste à présenter un état des lieux des termes et des initiatives de collecte de l'information géographique volontaire. Un deuxième aspect, quant à lui, s'intéresse à l'usage de l'information géographique volontaire dans les agences nationales de cartographie. Cette section résume donc les travaux de trois publications suivantes:

- See, L., P. Mooney, G. Foody, L. Bastin, A. Comber, J. Estima, S. Fritz, N. Kerle, B. Jiang, M. Laakso, H.-Y. Liu, G. Milcinski, M. Niksic, M. Painho, A. Podor, A.-M. Olteanu-Raimond and M. Rutzing (2016) Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information, ISPRS IJGI, vol. 5, n. 5, pp. 55.
- Olteanu-Raimond, A-M, Glen Hart, Giles M. Foody, Guillaume Touya, Tobias Kellenberger, and Demetris Demetriou. 2017. 'The Scale of VGI in Map Production: A Perspective on European National Mapping Agencies: VGI and the National Mapping Agencies'. *Transactions in GIS*.
- Olteanu-Raimond, A-M, Laakso, Mari, Antoniou, Vyron, Cidália Costa Fonte, Fonseca, Alexandra, Magdalena Grus, Harding, Jenny, Kellenberger, Tobias, Minghini, Marco, and Andriani Skopeliti. 2017. 'VGI in National Mapping Agencies: Experiences and Recommendations'. In *Mapping and the Citizen Sensor*, 299–326. Ubiquity Press.

2.2.1 Terminologies liée à l'information géographique volontaire

Dans un premier temps, nous nous sommes intéressés aux termes employés, à leur définition et à leur contexte d'utilisation. A partir d'une liste de termes, nous avons sélectionné une liste d'articles qui nous ont paru intéressants à étudier. Pour cela différents critères ont été identifiés tels que des articles de journaux et actes de conférences, les articles où apparaissent la première fois les termes, etc. Chaque article sélectionné a systématiquement été analysé par deux auteurs de l'article (See et al., 2016).

Suite à ce travail d'analyse, les différents termes ont été classifiés en fonction de leurs principales caractéristiques. Une première distinction a été faite entre les termes faisant référence à l'information collectée (e.g. IGV, AGI) et ceux faisant référence aux processus ou mécanismes qui peuvent aboutir à une collecte d'information (e.g. Citizen Science, Crowdsourcing). Les termes visant l'information collectée sont catégorisés selon deux axes (voir Figure 2).

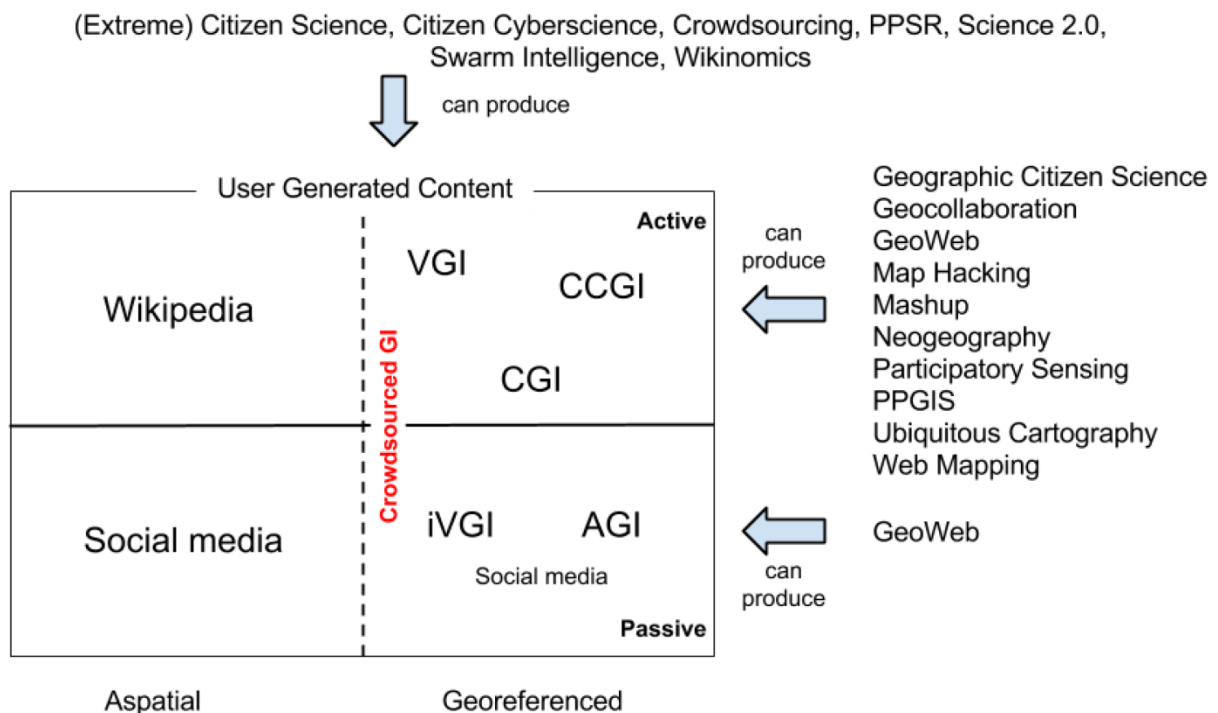


Figure 2. Catégorisation des termes concernant l'information produite par la foule où AGI: Ambient Geographic Information; CCGI: Citizen-contributed, Geographic Information or Collaboratively Contributed Geographic Information; CGI: Contributed Geographic Information; PPGIS: Public Participation in Geographic Information Systems; PPSR: Public Participation in Scientific Research; iVGI, VGI: Involuntary, Volunteered Geographic Information; issue de (See et al., 2016).

Un premier axe décrit si l'information collectée est géographique ou non géographique. Ainsi, même si une partie de l'information issue de Wikipédia et des réseaux sociaux peut-être géographique, la majorité de cette information générée par les utilisateurs n'est pas géographique. Cependant, dans la littérature, l'information provenant des médias sociaux a été utilisée comme une source d'information géographique issue de la foule. On peut citer, par exemple, l'utilisation de photographies provenant de la plateforme Flickr pour évaluer la classification de données Corinne Land Cover (CLC) (Yan et al., 2019) ou l'utilisation de tweets provenant de la plateforme Twitter pour déterminer l'épicentre ou le ressentiment de tremblements de terre (Crooks et al., 2013; Sakaki et al., 2010). Concernant l'information géographique, nous pouvons citer les termes VGI, CCGI, iVGI. Un deuxième axe décrit la manière dont la collecte d'information a été réalisée et distingue deux types : volontaire et involontaire. L'information dite volontaire (e.g. VGI) se réfère au fait que les contributeurs font la démarche de produire et partager la donnée (e.g. traces de randonnées) alors que l'information dite involontaire (iVGI) fait référence au fait que l'information est collectée dans un

autre but et que le contributeur produit l'information à travers l'usage d'une fonctionnalité ou d'un dispositif (e.g. traces issues de téléphones portables).

2.2.2 Initiatives pour la collecte de l'information géographique volontaire

Afin d'évaluer la manière dont l'IGV est produite et dans quel but, nous avons analysé différentes initiatives impliquant la collecte de l'IGV. Pour cela, le point de départ a été le site web VGI-net⁸, créé en 2011 et maintenu par les chercheurs de l'Université de Californie, Santa Barbara, l'Université d'État de l'Ohio et l'Université de Washington pour recenser les différentes initiatives collaboratives. Le site web n'étant pas mis à jour régulièrement, la première tâche a été d'éliminer les sites web recensés qui n'étaient plus en activité, et d'ajouter de nouveaux sites non recensés. Ce travail a donné lieu à une centaine de sites web qui ont été examinés. Cet examen ne se veut pas exhaustif, car les sites web et les plateformes changent constamment. Il s'agissait d'un état des lieux pour une période donnée destiné à fournir un échantillon suffisamment important pour permettre de tirer des conclusions générales sur la connaissance de l'IGV provenant de la foule.

Les sites web ont été analysés selon sept critères : le thème, la nature et le type de l'IGV, l'expertise et la formation, l'accessibilité aux données et aux métadonnées, la qualité et l'utilisabilité, les informations concernant les participants, les incitations à la participation. Je résume brièvement les conclusions tirées selon les sept critères ; pour plus d'informations, voir (See et al., 2016).

Thème. Au plus haut niveau, les sites web et les applications mobiles peuvent être divisés en trois grands types : (i) ceux qui permettent aux utilisateurs de créer et de partager une carte ; (ii) ceux qui collectent de l'information géographique; et (iii) les sites web de partage de données au niveau européen fournis par des experts mais qui peuvent inclure des données collectées par des citoyens. Parmi les cent sites web analysés, 12 étaient du premier type et quatre du troisième type. Ainsi, la majorité des sites web/applications concernent la collecte de l'IGV. La catégorie de site web la plus fréquente concerne l'écologie (e.g. l'identification des espèces). Ce nombre élevé n'est pas surprenant compte tenu des pratiques des sciences citoyennes dans ce domaine, remontant à des décennies et même des siècles avant l'apparition de l'internet. D'autres sites web (plus de cinq) concernent le suivi environnemental, les médias sociaux où la localisation joue un rôle essentiel dans le lien de connexion virtuel entre les utilisateurs (e.g. les sites de mise en relation des personnes en se basant sur leur proximité spatiale), les sites d'intérêt ou de voyages avec partage de photos, vidéos et récits de voyage géolocalisés, les sites pour la saisie des données topographiques (e.g. OSM), les sites de stations météorologiques amateurs offrant des informations sur la météo, l'état de la neige et le signalement des avalanches. Une catégorie qui nous a semblé sous-représentée dans notre échantillon de cent sites web est la cartographie de crise. Ceci est dû au fait que les sites web ont tendance à apparaître pendant les événements (e.g. disparition du vol de Malaysia Airlines MH370 où plus de 2 millions de personnes se sont mobilisées à la recherche de l'avion disparu en analysant des images satellites) et à disparaître après ceux-ci, ou parce que les contributeurs sont souvent recrutés sur le terrain et la saisie de l'information se fait en interne dans les organisations. Cependant, il existe au moins trois sites permanents qui méritent d'être mentionnés, à savoir Ushahidi⁹ Tomnod¹⁰, et le volet humanitaire d'OSM¹¹.

Type. Les différents types d'information géographique volontaire ou involontaire sont catégorisés selon deux axes (voir la Figure 3). Le premier axe vise à savoir si un type d'information est traditionnellement produit par une institution telle que les ANC ou s'il s'agit d'un type d'information qui peut être complémentaire (i.e. Framework data vs Non-Framework data). Le deuxième axe,

⁸ <http://vgi.spatial.ucsb.edu/>

⁹ <http://www.ushahidi.com> : plateforme permettant aux habitants des zones touchées par un événement catastrophique de télécharger et de voir les informations géoréférencées en ligne

¹⁰ <http://www.tomnod.com> : plateforme citoyenne pour la cartographie de dommages

¹¹ <http://hotsom.org> : cartographie ouverte pour l'aide humanitaire

quant à lui, sépare le type d'information en fonction du caractère volontaire ou involontaire de la contribution (*i.e.* passive vs active). Nous avons remarqué une forte complémentarité avec les données d'autorité et une large palette des sites collectant les informations météorologiques.

Expertise et formation. Ce critère s'intéresse aux sites web/applications du point de vue de l'expertise nécessaire aux utilisateurs pour y participer et de l'existence de formations ou de manuels d'utilisation. Ce critère concerne uniquement les sites web/applications dont la contribution est volontaire (*i.e.* partie droite de la Figure 3). Nous avons classé les sites selon trois niveaux de difficulté. Le niveau de difficulté faible regroupe les sites web/applications qui nécessitent très peu d'expertise pour pouvoir participer, à l'exception de l'usage d'Internet ou d'un téléphone portable (*e.g.* saisie d'un formulaire, téléchargement de photos, commentaires). Ils ne disposent ni de guides d'utilisation, ni de formations. Le deuxième niveau de difficulté concerne les sites web/applications dont la participation nécessite d'être familier avec un domaine (*e.g.* météo, astrologie, identification d'objets sur une image satellite). Ces sites proposent un matériel de formation, des instructions, des vidéos et/ou des FAQ, que les participants peuvent consulter. Bien que l'expertise requise soit minime, la participation exige, tout de même, un certain effort d'apprentissage de la part des participants.

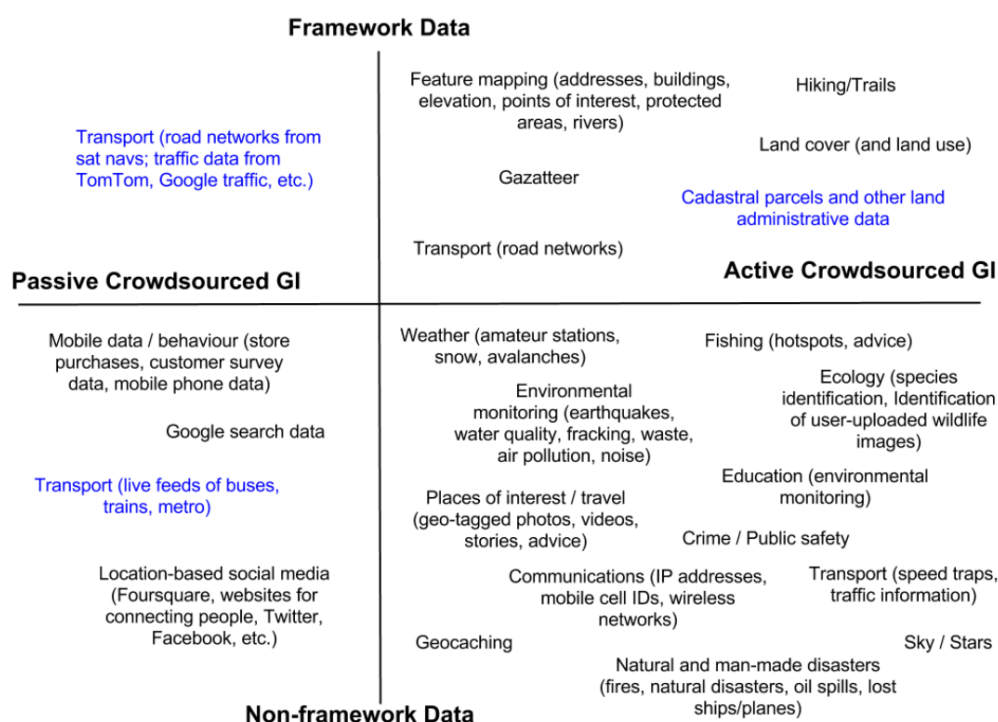


Figure 3. Types d'information géographique volontaire et involontaire provenant des initiatives citoyennes. Les types d'information non présents dans les cent sites analysés et issus de l'analyse des publications scientifiques ont été ajoutés en bleu ; issue de (See et al., 2016).

Enfin, le troisième niveau de difficulté regroupe les initiatives destinées à la collecte d'information concernant les pratiques sportives, l'écologie et la météo. Pour les pratiques sportives, la familiarité avec l'utilisation d'un système de positionnement global (GPS) est requise. La présence de manuels de formation de qualité est variable d'un site à un autre. Dans le cas des initiatives d'ordre écologique, une plus grande expertise est nécessaire et les participants doivent suivre un protocole strict de collecte de données. Dans quelques cas, la présence physique à une séance de formation est nécessaire. Enfin, pour les initiatives dans le domaine de la météo, les stations météorologiques d'amateurs exigent des connaissances sur l'installation, qui doit être conforme à certaines règles afin d'assurer la qualité de l'information collectée. En fonction de l'usage des données envisagé (*e.g.* recherche ou à d'autres fins faisant autorité), des formations, des manuels d'utilisation ou même des modèles de prévision étaient disponibles.

Disponibilité de l'information collectée et de métadonnées. La disponibilité varie selon les initiatives, allant d'indisponible (*i.e.* utilisée uniquement en interne), à disponible uniquement aux participants, ou aux participants enregistrés et connectés, ou ouverte à tous. Dans le cadre de ces différents niveaux de disponibilité, l'information collectée était disponible pour consultation sur une interface cartographique, en téléchargement dans divers formats (*e.g.* CSV, KML) ou via une API (Interface de Programmation d'Application). Certaines initiatives publient l'information collectée soit sous forme brute (*e.g.* absence de prétraitement), soit sous forme agrégée. Nous avons remarqué que les initiatives proposant une connexion via l'API sont celles ayant un volume de données important et une demande forte d'accès aux données (*e.g.* media sociaux, activités sportives). Les métadonnées au sens de la norme ISO 19115 ne sont pas mentionnées. En revanche, des métadonnées au sens de la description des données peuvent être disponibles. Le détail et la qualité varient d'une initiative à une autre. Nous avons remarqué que les sites disposant de protocoles de collecte proposaient également de métadonnées détaillées et les initiatives dont la difficulté de participation étaient élevée proposaient des métadonnées détaillées pour chaque jeu de données publié sur le site web/ l'application.

Qualité et utilisabilité. Pour la majorité des initiatives, il est difficile de conclure s'il existe un contrôle de qualité car ce dernier peut ne pas être visible pour l'utilisateur. Pour la majorité des sites web/ applications analysées, nous concluons que l'information géographique collectée ne semble pas faire l'objet d'un contrôle de qualité. Cependant, nous avons identifié certains sites web/ applications qui appliquent des méthodes de vérification automatique (*e.g.* valeurs en dehors des plages de valeurs autorisées), un retour sous forme de commentaire de la part des administrateurs, ou l'examen d'un expert. Un contrôle qualité est présent pour la plupart des sites écologiques météorologiques ou bien pour les sites de cartographie collaborative tels qu'OSM et Google MapMaker¹².

Informations concernant les participants. Les sites web/ applications peuvent être classés en trois catégories en fonction des informations d'ordre privé qu'ils obtiennent sur les contributeurs : (i) aucun enregistrement requis, donc pas de lien entre l'information collectée et le participant; (ii) l'inscription est requise mais seuls le nom et l'adresse électronique sont demandés; (iii) l'inscription est requise avec des informations personnelles recueillies telles que l'adresse, l'organisation, l'âge, le niveau d'expertise, la motivation, la connexion via un autre compte existant (*e.g.* Facebook, Google). La majorité des sites web/ applications étudiés appartenaient aux deux premières catégories, ce qui implique que très peu d'analyses sur le profil des contributeurs peuvent être réalisées.

Incitations à la participation. Deux incitations ont été identifiées : (i) participer à une bonne cause ou contribuer au bien commun, ce qui implique souvent un flux d'informations à sens unique (*e.g.* la cartographie des dommages) et (ii) un intérêt personnel (*e.g.* recevoir des informations trafic en temps réel, améliorer le quotidien) qui impliquent souvent un flux d'informations à double sens. Dans les deux cas, mais pour un nombre restreint de sites examinés, nous avons remarqué des incitations supplémentaires intégrées dans le processus de collecte (*e.g.* les forums de discussion, la ludification par le biais des tableaux de bord et des classements), la reconnaissance des efforts par des niveaux de réussite et l'interaction avec des experts. Les initiatives qui nous semblaient avoir le moins de succès, étant donné le nombre de contributions récentes, sont celles qui n'offrent que le premier type d'incitation. Des exceptions évidentes sont OSM ou Google MapMaker, qui ont connu un grand succès ; c'est toujours le cas pour OSM.

2.2.3 Usages de l'information géographique volontaire au sein des agences nationales de cartographie

Dans la continuité des travaux présentés en Section 2.2.2, j'ai initié et piloté un groupe de travail dans le cadre de l'Action COST *Mapping and Citizen Sensors* ayant comme objectifs de dresser un

¹² Il convient de noter que le site Google MapMaker a été fermé le 31 mars 2017

état des lieux concernant l'usage de l'IGV au sein des ANC et d'identifier les barrières quant à l'usage de l'IGV dans des chaînes de production dites d'autorité.

Pour répondre à ces objectifs, nous avons d'abord réalisé une enquête en utilisant le service Survey Monkey. L'enquête a été distribuée auprès de 41 organisations en Europe et au Groenland via les communautés EuroSDR et EuroGeographics pendant la période Décembre 2013 -Février 2014. Pour compléter cette enquête, et notamment identifier les barrières quant à l'usage de l'IGV par les ANC, nous avons organisé deux ateliers dans les locaux de l'IGN-France en mars 2014 et Swisstopo en février 2015. Notons qu'au deuxième atelier, des chercheurs travaillant sur l'IGV ont été invités.

Au total, 39 organisations ont répondu à l'invitation, mais seules 23 réponses ont été complètes et utilisables. La discussion ci-dessous est basée sur les réponses de ces 23 organisations, provenant de 19 pays. Il convient de noter que dans certains cas, comme pour l'Allemagne, les réponses provenaient à la fois d'agences nationales et régionales. En outre, pour le Royaume-Uni, les organisations participantes pour la Grande-Bretagne et l'Irlande du Nord ont été traitées comme des organisations nationales plutôt que régionales.

L'enquête visait à obtenir des informations organisées autour de trois aspects: (1) le rôle de l'organisation, (2) les spécifications de production, de mise à jour et de contrôle qualité et (3) l'engagement actuel et futur avec l'IGV, y compris le contrôle de la qualité.

Le rôle de l'organisation. La grande majorité des ANC ont comme responsabilité principale la production et la cartographie de données topographiques, mais d'autres responsabilités ont également été notées (*e.g.* hydrologie, cadastre, météorologie).

Spécifications de production, de mise à jour et de contrôle qualité. L'enquête a mis en évidence une différence considérable entre les ANC au sujet de l'échelle cartographique et la précision des données produites. Les échelles cartographiques sont de 1:10 000 ou moins et les données topographiques produites ont une précision pour la plupart des cas inférieure à 5m. Ceci confirme que le niveau de précision est bien supérieur à celui nécessaire aux besoins cartographiques. Plus important encore, cela veut dire que les contributeurs volontaires pourraient fournir des données avec une précision géométrique suffisante pour répondre aux besoins cartographiques.

Les politiques de mise à jour, quant à elles, montrent que les politiques cycliques sont plus largement utilisées que les politiques continues. Plus précisément, 11 agences ont adopté une politique cyclique, 6 une politique continue et 2 ont à la fois des mises à jour cycliques et continues, bien que leur principale politique de mise à jour soit continue. De manière inattendue, l'enquête a montré que 4 des organismes n'avaient pas encore établi de politiques, ce qui montre qu'elles sont relativement nouvelles. Dans le contexte de l'utilisation de l'IGV, on s'attend à ce que les contributeurs soient capables de fournir des informations sur les changements pour les thèmes dont la révision est continue (*e.g.* routier), car l'effet de la contribution est vu dans une période de temps courte. Il serait donc envisageable de proposer des initiatives pour les thèmes dont la mise à jour est cyclique mais cela devrait être réalisé dans une activité encadrée et dirigée. Par exemple, les volontaires sont encouragés à se concentrer sur une tâche ou un type de données à un moment donné.

L'enquête a également mis en évidence une pratique communautaire que nous pouvons appeler d'autorité. Dix-sept agences reçoivent des données terrain issues des partenaires (*e.g.* les pompiers, les mairies). Ces données sont accompagnées de métadonnées et soumises systématiquement à un contrôle qualité. La pratique de la vérification reflète d'une part la priorité des ANC de produire des données de haute qualité, homogènes et cohérentes, et d'autre part, le manque de méthodes permettant la vérification automatique et l'intégration de nouvelles fonctionnalités.

Engagement actuel et futur avec l'IGV, y compris le contrôle de la qualité. Historiquement, toutes les ANC ont utilisé une sorte d'IGV en répondant aux remontées d'erreurs ou d'incohérences cartographiques qui leur étaient signalées par le grand public ou par les acteurs publics. Toutefois, il est important de faire la différence entre ce processus *ad-hoc*, dans lequel l'ANC est essentiellement

un participant passif, et les initiatives dans lesquelles elles s'engagent activement, soit en proposant leur propre plateforme de collecte, soit en utilisant l'IGV existante. Ainsi, comme l'illustre la Figure 4, l'enquête et les ateliers ont indiqué que douze ANC sur vingt-trois utilisent l'IGV pour la détection des changements, la remontée d'alertes, la collecte de nouveaux contenus, les noms de lieux vernaculaires et l'interprétation de photos.

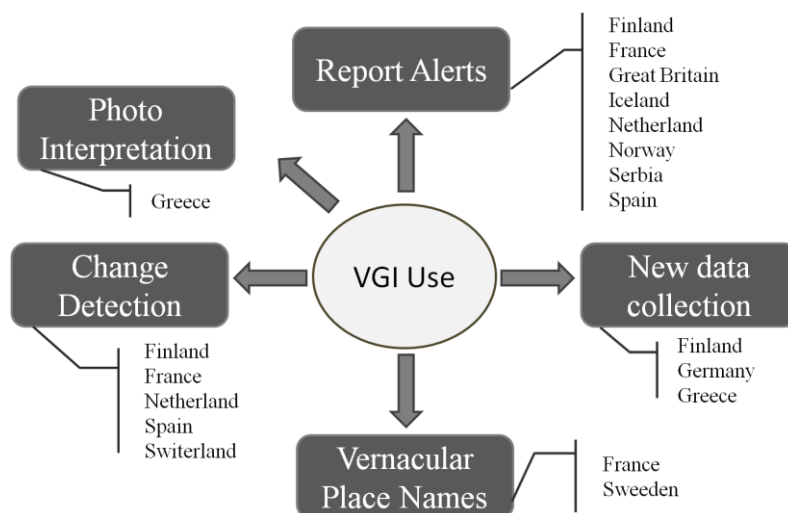


Figure 4. Thèmes impliquant l'usage de l'information géographique volontaire dans les ANC ; issue de (Olteanu-Raimond et al., 2017a)

La plupart d'entre elles ont proposé leurs propres plateformes de collecte de l'IGV, tandis que trois agences (*i.e.* Allemagne, Grèce, Serbie) ont fait appel à l'IGV issue d'OSM ou de Geonames¹³. Dans la plupart des cas, l'IGV collectée concernait les données topographiques produites traditionnellement par les ANC (voir Figure 5). Une caractéristique quasi universelle est que chaque ANC engagée activement dans la collecte de l'IGV a procédé à un contrôle qualité.

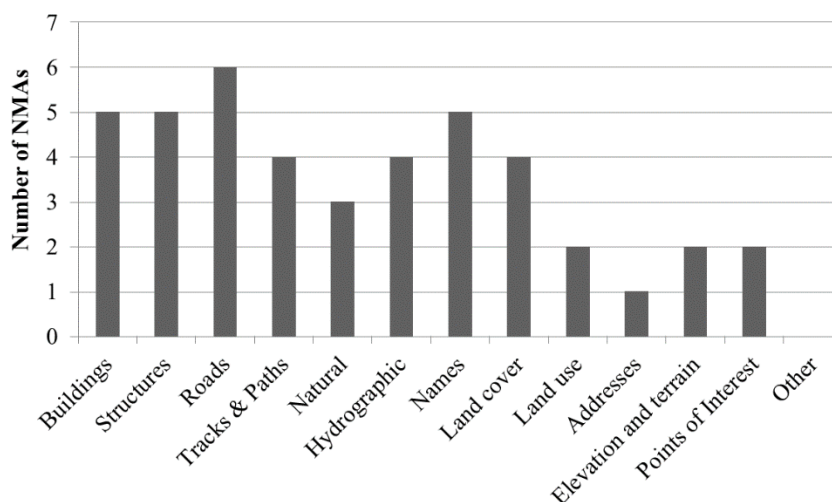


Figure 5 : Types d'information géographique volontaire utilisés par les ANC ; issue de (Olteanu-Raimond et al., 2017a)

Dans la plupart des cas, l'usage de l'IGV se limitait à la détection des changements et à la correction des erreurs. Seules deux agences utilisent l'IGV pour collecter des données qu'elles ne produisaient pas traditionnellement. Cette situation suggère que les ANC n'exploitent pas les nouvelles

¹³ <https://www.geonames.org/>

opportunités venant de l'IGV, en raison de différents obstacles, dont certains sont examinées ci-dessous.

Concernant la motivation des contributeurs, nous avons constaté qu'elle était d'ordre social (e.g. participer à améliorer la qualité des référentiels d'autorité) ou pour un intérêt particulier (c'est le cas des partenaires des ANC qui réutilisent l'information collectée). La motivation des contributeurs est incitée par le retour aux utilisateurs et la mise à jour des référentiels prenant en compte les remontées réalisées.

Neuf agences qui n'avaient pas utilisé l'IGV ont déclaré qu'elles ont l'intention de le faire. Enfin, celles qui n'ont pas envisagé cet usage ont évoqué des inquiétudes liées aux licences, à la qualité et à la fiabilité de l'IGV.

Ce premier résultat nous a mené à organiser les ateliers mentionnés ci-dessus afin de dresser une liste d'obstacles et d'identifier quels seraient les éléments qui permettraient une relation de confiance entre les communautés des ANC, les chercheurs et des contributeurs volontaires.

Cinq obstacles majeurs à l'utilisation de l'IGV ont été identifiés : (1) la qualité et la validation des données, (2) les questions juridiques de droits de données et responsabilité civile, (3) l'expertise et la motivation de la foule, (4) la motivation à long terme et (5) les craintes en matière d'emploi.

Enfin, une des conclusions des ateliers a été qu'une meilleure relation entre les ANC et la communauté de l'IGV pourrait être la clé du succès. Cela implique une relation de confiance, et une relation de type « gagnant-gagnant » où les attentes des agences et des communautés sont réciproquement satisfaites. La Figure 6 résume les attentes les plus importantes du point de vue des auteurs de (Olteanu-Raimond et al., 2017a).



Figure 6. Relations entre la communauté de l'IGV, les ANC et la communauté recherche ; issue de (Olteanu-Raimond et al., 2017a)

Pour la communauté de l'IGV, l'une des attentes les plus importantes est le retour d'information. Les agences doivent informer les contributeurs, montrer toutes les contributions, les valider et les intégrer le plus rapidement possible dans les données qui font autorité. À cette fin, une possibilité est que les ANC mettent à disposition les informations collectées via une initiative collaborative sous la forme d'une couche citoyenne pouvant être associée aux couches de données faisant autorité en permettant ainsi d'être utilisées aussi par les contributeurs et d'autres utilisateurs. Dans le même

temps, les ANC attendent de la communauté de l'IGV une motivation, une stabilité dans le temps, l'absence de contrefaçon, etc. Pour améliorer les relations avec la communauté VGI, les ANC attendent de la communauté de chercheurs de nouvelles méthodes de qualification et de contrôle de qualité, des processus fiables pour la résolution des conflits et l'intégration des données, de nouvelles techniques et des recommandations pour diminuer les obstacles identifiés.

2.3 Recommandations et mise en œuvre d'une plateforme collaborative pour la collecte de l'information géographique volontaire

Avant-propos

Cette section présente les travaux que j'ai réalisés en collaboration avec des collègues chercheurs impliqués dans l'action COST *Mapping and Citizen Sensors* et des collègues impliqués dans le projet LandSense. Les partenaires privilégiés ont été : IIASA, l'Université de Coimbra, de Milan, d'Athènes, Nottingham et de Maynooth, ainsi que les ANC de Finlande, Suisse, Pays-Bas, UK, Portugal. Deux aspects y sont traités : d'une part, les travaux qui visent à dresser des recommandations pour la collecte de l'information géographique volontaire d'une manière générale, d'autre part, les travaux décrivant un protocole de collecte et une plateforme communautaire pour le suivi de l'occupation du sol.

Cette section résume donc les travaux des quatre publications suivantes:

Mooney, P., Minghini, M., Laakso, M., Antoniou V., Olteanu-Raimond, A-M, and Skopeliti, A., 2016. 'Towards a Protocol for the Collection of VGI Vector Data'. *ISPRS International Journal of Geo-Information* 5 (11): 217. <https://doi.org/10.3390/ijgi5110217>.

Olteanu-Raimond, Ana-Maria, Laakso, Mari, Antoniou, Vyron, Cidália Costa Fonte, Fonseca, Alexandra, Magdalena Grus, Harding, Jenny, Kellenberger, Tobias, Minghini, Marco, and Andriani Skopeliti. 2017. 'VGI in National Mapping Agencies: Experiences and Recommendations'. In *Mapping and the Citizen Sensor*, 299–326. Ubiquity Press.

Minghini, M., Antoniou, V. Fonte, C., Estima, F., Olteanu-Raimond, A-M, See, L., Laakso, M et al. 2017. 'The Relevance of Protocols for VGI Collection'. In *Mapping and the Citizen Sensor*, 223–247. Ubiquity Press.

Olteanu-Raimond, A-M, Jolivet, L., Van Damme, M-D, Royer, T., Fraval L., See L., Sturn, T., Karner, M., Moorthy, I. and Fritz, S. 2018. 'An Experimental Framework for Integrating Citizen and Community Science into Land Cover, Land Use, and Land Change Detection Processes in a National Mapping Agency'. *Land* 7 (3): 103. <https://doi.org/10.3390/land7030103>.

2.3.1 Vers des protocoles de collecte de l'information géographique volontaire

Dans le cadre de l'Action COST *Mapping and Citizen Sensors*, nous avons proposé un protocole de collecte de l'information géographique vecteur qui vise à la fois la mise en place de nouvelles initiatives de collecte de l'IGV et un usage rétrospectif pour les initiatives collaboratives déjà existantes. Le protocole a donc été développé de manière bidirectionnelle : d'une part, nous avons étudié les pratiques de collecte et d'édition dans le cadre des projets collaboratifs existants, d'autre part, nous avons analysé les pratiques dites traditionnelles mise en œuvre par les autorités compétentes, telles que les ANC. Ainsi, notre protocole se situe à l'intersection de deux pratiques opposées pour la collecte et l'édition de l'information géographique. Il vise à augmenter la motivation des contributeurs, tout en répondant aux critères rigoureux des ANC pour produire une information géographique de qualité.

L'hypothèse de départ a été qu'un protocole de collecte doit suivre des objectifs du haut niveau (e.g. interopérabilité, adaptabilité, etc.) et de contenu (e.g. décrire comment faire la collecte de données, répondre aux questions potentielles des contributeurs, décrire l'information géographique

collectée). En partant de cette hypothèse, nous nous sommes focalisés sur trois aspects : le modèle de données, la méthode de collecte des données et les caractéristiques de l'IGV collectée.

Le modèle de données. Cet aspect consiste à décrire d'une manière précise l'objectif du projet collaboratif et le type d'information que l'on souhaite collecter. Dans la mesure du possible, le modèle de données doit être flexible afin de laisser place à l'initiative des contributeurs pour enrichir le modèle de données en ajoutant un attribut ou un type d'objet qui n'est pas prévu. Le protocole de données distingue trois types d'information qui devraient être détaillés afin d'éviter l'incomplétude et l'hétérogénéité de l'IGV collectée: le type de géométrie (*e.g.* point, ligne, surface, géométrie simple ou multiple ; rivière représentée par un linéaire ou une surface), les attributs thématiques (*e.g.* noms, nature, nombre de voies) et les règles de saisies (*e.g.* saisir l'axe de la chaussée pour une route départementale et l'axe de la chaussée pour les deux sens d'une autoroute). Concernant la nature d'un objet, le protocole de données recommande l'utilisation des taxonomies ou des listes de natures prédéfinies. Mettre à disposition des contributeurs des exemples est important afin d'éviter toute mauvaise interprétation. Le défi est de trouver un équilibre entre un manuel d'utilisation qui décourage le contributeur et une explication simpliste qui négligerait certains aspects qui peuvent impacter la qualité de l'information collectée.

Méthode de collecte des données. Cet aspect vise à traiter les méthodes de collecte de données. Nous nous sommes focalisés sur trois méthodes : la saisie à partir de fonds cartographiques, la collecte sur le terrain et l'import de l'IGV. Les recommandations ont porté sur les outils de collecte ou d'import (*e.g.* paramétrages, usages, l'échelle de saisie). Pour chaque méthode, nous avons listé des bonnes pratiques à suivre. Le lecteur est invité à voir Mooney et al. (2016) pour plus de détail.

Les caractéristiques de l'IGV collectée. Ce troisième aspect, quant à lui, s'intéresse à la description de l'IGV. Certaines descriptions sont obligatoires (*e.g.* système de projection, source des données, licence), d'autres ont un caractère optionnel (*e.g.* respecter le niveau de détail demandé, remplir les métadonnées, donner des éléments de qualité). Les descriptions obligatoires sont celles qui, en général, peuvent diminuer la motivation des contributeurs. Notre recommandation pour les concepteurs d'un projet de collecte de l'IGV a été d'intégrer dans les outils de collecte des fonctionnalités permettant l'enregistrement automatique des métadonnées (*e.g.* l'échelle de saisie, la cadence d'enregistrement d'un capteur GPS, la précision du GPS, l'horodatage). Demander aux contributeurs des informations sur la qualité de l'information collectée est une tâche difficile qui n'est pas innée pour tout profil de contributeur. Cependant, le contributeur devrait être encouragé à donner des informations utiles telles que la non mise à jour ou la faible résolution du fond cartographique utilisé, le dysfonctionnement de l'outil d'ancrage topologique, etc. Une liste des métadonnées souhaitées a été identifiée pour chaque méthode de collecte décrite ci-dessus.

Le protocole se veut générique afin que les différents types de contributeurs tels que les producteurs officiels d'information géographique, les associations publiques ou privées, les acteurs publics, les organisations non gouvernementales (ONG) ou la communauté des chercheurs puissent l'implémenter.

Afin de faciliter la prise en main du protocole, nous avons proposé une méthodologie composée de cinq étapes : l'initialisation, la collecte proprement-dite, l'auto-contrôle de qualité, la soumission de l'information collectée et le retour aux utilisateurs. Des recommandations ont été décrites pour chaque étape et instanciées à travers trois cas d'étude fictifs. Ce protocole a été par la suite étendu pour la collecte de photos géo-localisées qui peuvent être utilisées dans la production ou la validation des données caractérisant l'occupation et l'usage du sol (Minghini et al., 2017). La collecte des photos géo-localisées dans ce contexte nécessite que le contributeur soit guidé étape par étape dans sa tâche et que les outils de collecte enregistrent des paramètres qui permettent de reconstituer le paysage (*e.g.* orientation de la prise de vue) et d'estimer la précision de la prise de vue. Contrairement à la collecte de l'IGV en mode vecteur qui doit être exhaustive, la collecte de

photos géo-localisées, étant un travail réalisé sur le terrain, nécessite de mettre en place des techniques d'échantillonnage (Olofsson et al., 2014; Stehman et Foody, 2019).

Dans Mooney *et al.*, (2016) et (Minghini et al., 2017), nous avons décrit un protocole générique et proposé une implémentation s'appuyant sur des cas fictifs. Pour aller plus loin, nous nous sommes intéressés à une adaptation du protocole de collecte décrit ci-dessous dans le cadre des projets collaboratifs applicables aux ANC qui intégrerait davantage les spécificités de production de ces dernières. Ainsi, dans un premier temps, avec des collègues de différents ANC d'Europe, nous avons proposé une extension du protocole de collecte sous forme de recommandations. En se basant sur les exigences des ANC en termes de sécurité et de qualité, mais également sur les expériences de certaines ANC à organiser et gérer leur propre plateforme de collecte de données, cela nous a semblé une bonne direction à prendre. Les recommandations sont groupées autour de six éléments : le modèle de données et les objets, l'interface, la motivation, l'identification, les licences, et le contrôle de la qualité. La Figure 7 illustre le workflow proposé pour la collecte de l'IGV dans le cadre d'un projet collaboratif mené par une ANC où ces six éléments sont marqués d'un astérisque (*). Les flèches vertes et roses représentent respectivement les tâches effectuées par les ANC (NMA dans la figure) et les contributeurs.

Notre hypothèse a été qu'une plateforme collaborative pourrait être une initiative réussie si elle est dédiée à la fois aux contributeurs et aux utilisateurs. Ainsi, une des recommandations est d'impliquer les citoyens, les groupes de citoyens ayant les mêmes intérêts, et des groupes d'organismes publics et gouvernementaux, ainsi que les acteurs de l'Education Nationale. L'engagement avec différents organismes publics et avec le secteur éducatif augmentera la participation des citoyens puisque ces organismes sont proches des citoyens et peuvent investir dans l'avenir en éduquant et en sensibilisant les jeunes générations à l'information géographique et à sa qualité. Ces engagements pourraient créer une motivation, assurer la durée et promouvoir l'IG de qualité tant pour les ANC que pour les contributeurs.

Pour une relation gagnant-gagnant, la plateforme doit offrir la possibilité aux contributeurs (citoyens, partenaires publics) de collecter des nouveaux types d'objets, qui ne sont pas collectés par les ANC, soit en raison de coûts nécessaires importants, soit parce qu'ils représentent des données topographiques non traditionnelles. Ces informations pourraient être utiles aux citoyens et aux divers services publics et agences gouvernementales. Les contributions doivent être faites via des interfaces conviviales qui mettent en œuvre un modèle de données adaptatif à partir de celui proposé par les ANC, via une identification sécurisée et des outils faciles à utiliser pour contribuer, gérer, visualiser et télécharger l'IGV ou les données d'autorité, en fonction de la licence de données de chaque ANC. Une valeur ajoutée des ANC est le contrôle de la qualité de l'IGV, qui peut être corrigée, validée et intégrée dans la plateforme collaborative. En fonction des spécifications, certaines IGV validées peuvent être intégrées dans le référentiel faisant autorité (le passage de Q-VGI à NMA dans la Figure 7), améliorant ainsi la qualité des données d'autorité. Le contrôle de la qualité pourrait être effectué de deux manières : continu à travers le partage d'opinions des contributeurs ou cyclique (mais dans une durée limitée) à partir des outils de contrôle développés par les ANC.

Un autre aspect important, orienté vers les citoyens, qui peut accroître la motivation est la ludification. Cependant, lors de la mise en œuvre de systèmes de récompenses (gamifiées ou réelles), il convient de prêter attention au fait que la qualité de données est beaucoup plus importante que la quantité, et cela doit être clairement expliqué au contributeur. Ainsi, une bonne pratique en matière de ludification consiste à éviter de donner des récompenses ou des prix basés seulement sur le nombre de contributions effectuées.

Le Tableau 1 résume les recommandations proposées et fournit une liste des avantages et inconvénients qui peuvent découler d'une telle plateforme collaborative par rapport aux six éléments identifiés.

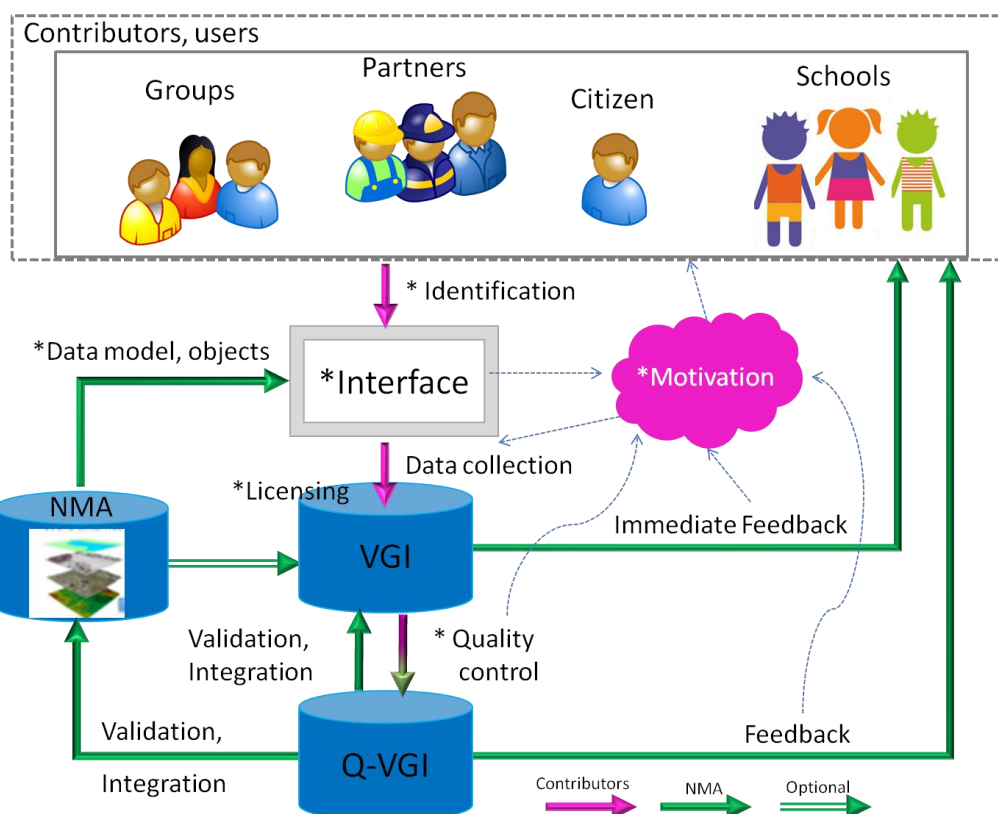


Figure 7. Proposition d'un workflow pour la collecte de l'IGV au sein d'une ANC ; d'après (Olteanu-Raimond et al., 2017b)

Tableau 1. Résumé des opportunités, inconvénients et recommandions pour une plateforme collaborative en fonctions des éléments : le modèle de données, l'interface, la motivation des contributeurs, la connexion, la licence et le contrôle qualité.

Éléments clés de la plateforme collaborative	Opportunités, SI intégration de l'IGV	Inconvénients, SI intégration de l'IGV	Recommandations
Modèle de données	- Nouveaux objets ; - Amélioration de l'exactitude ; - Enrichissement thématique ; - ANC peut avoir un rôle moteur dans le développement de standards pour la collecte de l'IGV.	- Volume de données trop important à gérer ; - Thématique trop hétérogène ; - Incomplétude (couverture et thématique).	- Proposer une couche citoyenne ; - Proposer un protocole d'édition pour les différents types d'objets à collecter ; - Proposer un modèle de données adaptative pour les nouveaux types d'objets ; - Proposer un forum et une aide en ligne ; - Faire un retour aux contributeurs ; - Développer des outils conviviaux et faciles à utiliser pour l'import, l'édition et le téléchargement ; - Bien distinguer l'IGV qualifiée par l'ANC.

Interface		- Faible nombre de contributeurs ou de groupes de contributeurs; - Démotivation des contributeurs.	- Facile à utiliser ; - Conviviale ; - Rapide ; - Fournir un guidage de type étape par étape pour faciliter la collecte de l'IGV en suivant le protocole de données proposé.
Motivation	- Recruter et maintenir un large nombre de contributeurs ; - Améliorer la qualité de l'IGV	- Contributeurs déçus qui ne reviennent pas ; - Non usage de l'IGV dans la production de données d'autorité ; - Sentiment d'injustice si il n'y a pas de relation « gagnant-gagnant ».	- Lancer des campagnes limitées dans le temps ; - Introduire des éléments de ludification ; - Impliquer et sensibiliser la jeune génération en introduisant des notions telles que la responsabilité citoyenne, la qualité de données dans la formation des écoles (primaires, secondaires, universitaires) ; - Réaliser des partenariats avec les pouvoirs publics à travers une relation « gagnant-gagnant ».
Enregistrement / Identification	- Les contributeurs peuvent être contactés pour leur envoyer un retour ; - Possibilité d'analyser le comportement des contributeurs et d'en dériver des indicateurs de qualité de l'IGV.	- Verrous lié à la vie privée ; - Données fournies fausses ou incorrectes ; - Si trop d'information demandée lors de l'enregistrement, démotivation des contributeurs ; - Réticence à l'enregistrement et à l'identification.	- Demander un enregistrement minimum (ex. un mail valide) ; - Possibilité d'envoyer un retour anonyme aux contributeurs ; - Proposition de trois niveaux d'enregistrement : forte (e.g. nom complet, code postal, profession et institution, numéro de téléphone, numéro de passeport/carte d'identité), légère (email valide, profession groupe d'âge, genre) et faible (e.g. email et mot de passe).
Licence		- Licences incohérentes si import de l'IGV depuis une autre source d'information avec une licence différente.	- Les ANC devraient préparer des <i>Termes et conditions</i> claires et rigoureuses et demander la validation avant la contribution ; - Les contributeurs devraient céder à l'ANC tous les droits sur les données ; - Transparence sur l'usage de l'IGV. Informer les contributeurs qu'en contribuant, ils fournissent de l'IGV qui pourrait être intégrée dans les données d'autorité.
Contrôle de qualité	- Le contrôle de qualité (semi) automatique peut réduire le temps de contrôle manuel (qui est fait actuellement) ; - Organiser des campagnes	- Trop d'information est demandée aux contributeurs ; - Un nombre important de contributeurs ne garantit pas	Proposer un contrôle qualité à plusieurs niveaux : - Niveau 0 : Contrôle en temps-réel (au moment de la contribution) ; - Niveau 1 : Contrôle post-traitement (i.e. appliquer des méthodes de

	de validation avec les contributeurs pour compléter le contrôle (semi) automatique.	nécessairement la qualité de l'IGV.	qualification à l'IGV collectée) ; - Niveau 2 : Contrôle par les contributeurs ; - Niveau 3 : Validation finale en respectant une typologie de règles de qualité.
--	---	-------------------------------------	---

Pour une description détaillée des recommandations, j'invite le lecteur à se référer à (Olteanu-Raimond et al., 2017).

Ce type de plateforme basée sur les recommandations discutées dans cette section peut être mise en place de manière progressive à travers la proposition de projets de recherche pour les ANC ayant un service de recherche tel que l'IGN France ou l'Agence finlandaise de cartographie (NLS) ou bien à travers des projets pilotes.

2.3.2 Implémentation d'un protocole de collecte pour le suivi de l'occupation du sol

Le protocole de collecte et les recommandations n'ont pas fait l'objet d'une évaluation formelle. Dans cette section, je décris l'implémentation du protocole et des recommandations dans un exemple concret de collecte d'IGV pour la mise à jour de la base de données OCS-GE (Occupation du Sol à Grande Echelle). Cette initiative s'inscrit dans le cadre du projet LandSense, où l'IGN a mené un cas d'étude dans la Région Occitanie.

L'OCS-GE est une base de données vectorielle à l'échelle régionale produite récemment par l'IGN « à la demande » avec le co-financement de l'Etat et de la Région. Sa spécificité est la distinction entre la couverture (*i.e.* décrit les éléments structurant le paysage sans prendre en compte leur fonction) et l'usage (*i.e.* décrit les usages des éléments structurant le paysage). Ainsi, chaque objet géographique (polygone) est caractérisé par une classe de couverture (*e.g.* bâti) et une classe d'usage (*e.g.* résidentiel). La nomenclature d'usage utilisée est conforme à la nomenclature INSPIRE et suit une classification hiérarchique allant jusqu'aux quatre niveaux (INSPIRE, 2013). La base de données OCS-GE® est produite en suivant trois étapes : (1) dans un premier temps, l'ossature est construite à partir des éléments stables dans le temps tels que le réseau routier et ferré ; (2) ensuite, une classification semi-automatique est réalisée en utilisant différentes bases de données topographiques produites par l'IGN (*e.g.* BDTOPO®, BDAdresse®) ; (3) enfin, un travail de photo-interpretation est réalisé pour rectifier la géométrie et les classes de couverture et compléter les classes d'usage. L'OCS-GE® est millésimée (*i.e.* la date de validité de la donnée correspond à celle du référentiel image utilisé) et mise à jour d'une manière cyclique, tous les trois ans.

Dans la zone d'étude concernée par le projet LandSense, il existe actuellement deux millésimes : 2013 et 2016. Chaque millésime comporte une nomenclature de 14 classes de couverture et 17 classes d'usage. Cependant, toutes les combinaisons de classes d'usage et de couverture, c'est-à-dire 14×17 , ne sont pas possibles.

Dans ce contexte de production et étant donné les objectifs du projet LandSense, notre démarche de recherche a été réalisée selon trois étapes : l'analyse des besoins, la proposition d'un protocole adapté aux besoins et l'implémentation.

2.3.2.1 Analyse des besoins

La première étape a consisté à identifier les besoins de l'IGN et des utilisateurs finaux pour améliorer la production et la mise à jour de l'OCS-GE®. Le recueil des besoins a été réalisé à travers (1) des réunions d'information avec différents agents du département de production de l'IGN, (2) l'analyse de différentes études de besoins des utilisateurs concernant l'occupation du sol, réalisées par l'IGN lors du démarrage de la production de l'OCS-GE® et (3) l'organisation d'un workshop réunissant 24

participants représentant des Directions Départementales Territoriales, des agences d'urbanisme, des collectivités locales, la communauté OSM., etc. Ces besoins ont été classés par l'équipe projet en fonction de leur priorité, leur fréquence (le nombre d'acteurs remontant le même besoin) et des objectifs du projet LandSense. Ainsi, nous avons décidé dans le cadre de LandSense de nous focaliser sur les besoins suivants :

- Distinguer la classe d'usage US235 *Production secondaire, tertiaire et usage résidentiel*. En effet, les utilisateurs finaux ont exprimé le besoin d'avoir une classification plus fine qu'une classe mixte, US235 comme c'est le cas actuellement.
- Améliorer l'usage du sol des types de classe difficiles à mettre à jour par photo-interprétation, à savoir les zones agricoles (*i.e.* les distinguer par rapport aux zones non-agricoles) et les carrières (*i.e.* déterminer si les carrières sont en activité ou non).
- Mettre à jour l'usage de la classe US6.1 *Zones en transition*. Il s'agit des zones avec des travaux en cours au moment de la production d'un millésime donné. Si les travaux sont finalisés, la classe d'usage doit être mise à jour pour le millésime suivant.
- Apporter des éléments de changement qui permettraient la réduction du temps destiné au suivi de changement.

2.3.2.2 Proposition du protocole de collecte de données

La deuxième étape a consisté à définir le protocole de collecte de données afin de répondre aux objectifs choisis. Le protocole de collecte comporte plusieurs aspects : les types de contribution et les campagnes, l'architecture et les rôles d'une plateforme collaborative, les fonctionnalités des outils et la mise en place des campagnes. Pour chaque aspect, différents concepts et mécanismes de collaboration ont été proposés. Ils sont décrits brièvement ci-dessous.

Tout d'abord nous avons identifié les tâches collaboratives à réaliser. Il s'agit de cinq tâches : (1) éditer un objet géographique (*e.g.* modifier sa géométrie ou l'information attributaire, supprimer ou ajouter un nouvel objet), (2) ajouter de nouvelles informations définissant l'objet géographique (*e.g.* ajouter et remplir un nouvel attribut qui n'existe pas dans la base de données initiale), (3) envoyer une alerte (*e.g.* remonter une information qui concerne une erreur dans la donnée) (4) valider une information (*e.g.* confirmer ou infirmer un changement détecté automatiquement) et (5) prendre une photo selon la direction indiquée. En ce qui concerne les **contributions** (*i.e.* information géographique volontaire fournie par un contributeur), nous avons défini deux concepts : les **contributions terrain (ou *in-situ*) et bureautique**. Le premier type constitue l'ensemble de contributions réalisées sur le terrain en utilisant un terminal mobile. Le deuxième type, quant à lui, englobe les contributions réalisées au bureau en utilisant un terminal bureautique. Pour les contributions terrain, nous avons proposé le concept de **lieu à visiter**. Ce dernier, proche du concept proposé dans (Laso Bayas et al., 2016), consiste à définir un lieu nécessitant une information supplémentaire et à l'afficher sur la carte afin que les contributeurs puissent le sélectionner, le visiter et remplir les informations demandées.

Nous avons proposé que la collecte se fasse dans le cadre d'une campagne et avons distingué trois types de campagnes. La **campagne opportuniste** consiste à définir préalablement une période de campagne, les types de contribution et de laisser les contributeurs choisir le moment de la contribution et le type de contribution. Elle implique la totalité de l'OCS-GE sur la zone d'étude. Un deuxième type est la **campagne guidée** qui donne la possibilité à l'organisateur de campagne de spécifier les tâches, de sélectionner un ou plusieurs types d'objet et de redéfinir des zones d'études plus spécifiques. Enfin, la **campagne « flash »** consiste à organiser des campagnes limitées dans l'espace et dans le temps afin de recueillir rapidement des informations. Ces campagnes peuvent se concentrer sur un thème spécifique ou sur la totalité des thèmes. Notons que les campagnes opportunistes et flash s'approchent du concept de carto-party de la communauté OSM.

Enfin, nous avons identifié les types d'utilisateur de la plateforme collaborative et le rôle de chacun. Trois profils d'utilisateur sont définis, à savoir l'Administrateur, l'Expert et le Contributeur. Chaque profil a un rôle spécifique dans le processus de collaboration. Par exemple, l'administrateur a pour rôle de définir, de lancer et de fermer les campagnes en gérant l'IGV collectée sur la plateforme. L'expert définit et lance une campagne guidée en identifiant une zone, les lieux à visiter, ou en gérant des services (e.g. un service de détection des changements). Le profil de l'expert est validé par un administrateur et est généralement réservé aux agents de l'entité qui organise la campagne ou partenaires de l'entité. Les rôles des contributeurs correspondent aux tâches dans le processus collaboratif. La Figure 8 illustre les rôles du profil *contributeur*. Les contributeurs sont volontaires et peuvent appartenir à différentes communautés, par exemple, les citoyens, les agents des institutions qui produisent de l'information géographique, les étudiants, les chercheurs, etc.

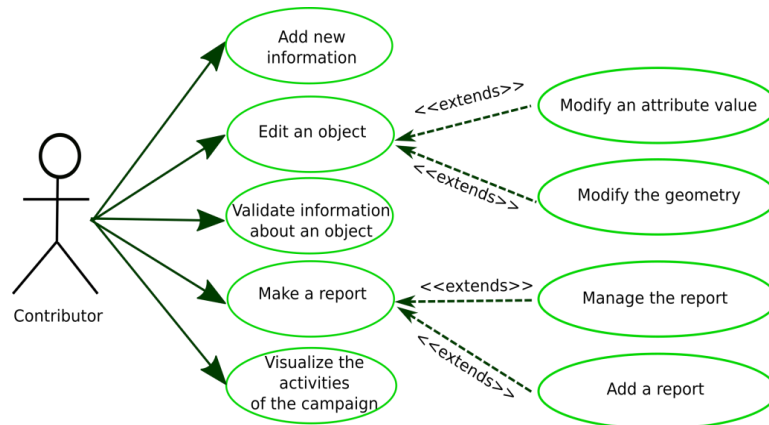


Figure 8. Les rôles du profil *Contributeur* correspondent aux tâches définies dans le processus collaboratif ; issue de (Olteanu-Raimond et al., 2018)

2.3.2.3 Implémentation de la plateforme PAYSAGES

La troisième étape a consisté à implémenter le protocole de collecte de données et a donné lieu à la plateforme PAYSAGES. Ce travail d'implémentation a été réalisé en étroite collaboration avec le département de développement de l'IGN, impliqué dans le projet LandSense, des collègues du laboratoire LASTIG et le partenaire LandSense, IASA. Dans le cadre de ce travail d'implémentation, j'ai assuré le suivi des développements en ayant, par ailleurs, le rôle de *product owner*.

Comme l'illustre la Figure 9, l'architecture de la plateforme PAYSAGES comporte un système d'authentification (Auth.), trois outils (PAYSAGES web, mobile et wiki), une API (REST), et une API (WFS-T) qui accède aux flux WMS du Géoportail.

Un compte d'utilisateur est nécessaire pour accéder à la plateforme PAYSAGES. Le contributeur peut se connecter avec un compte déjà existant (e.g. Google) ou en créer un. Dans ce dernier cas, comme suggéré dans (Olteanu-Raimond et al., 2017), deux types d'informations sont demandées aux contributeurs pour créer un compte : obligatoires (i.e. l'identifiant, le mot de passe et l'adresse électronique) et facultatives (i.e. la profession, l'âge, le sexe, l'institution). Etant donné que le système d'authentification fédéré était en cours de développement lorsque nous avons effectué les premières campagnes, la plateforme PAYSAGES a été mise en production sans le système d'authentification fédéré proposé dans LandSense (Matheus et Batič, 2016). Ceci implique que les contributeurs ont créé un compte. Sauf quelques exceptions, les contributeurs n'ont pas rempli les informations facultatives.

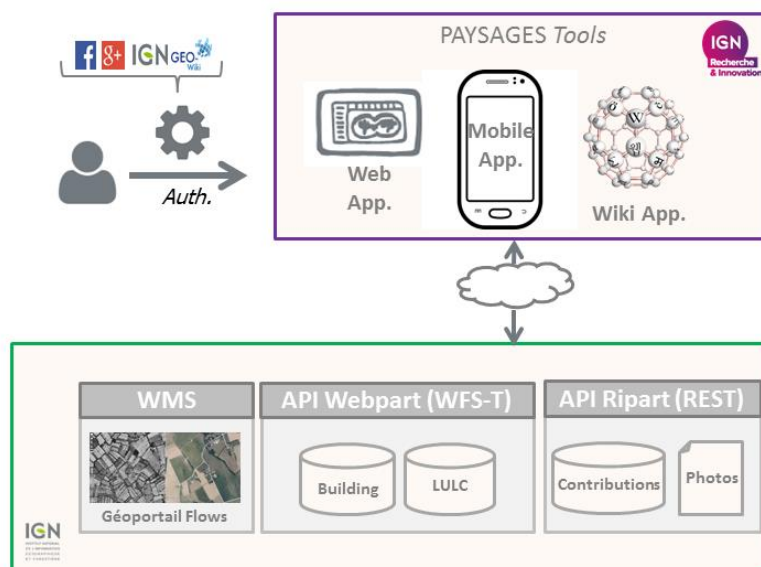


Figure 9. L'architecture de la plateforme collaborative PAYSAGES ; issue de (Olteanu-Raimond et al., 2018)

L'outil [PAYSAGES web](#) (voir la Figure 10, à droite) est une application serveur qui s'appuie sur les briques logiciel de l'outil développé par l'IGN, nommé [Espace collaboratif](#). De nouvelles fonctionnalités implémentant les besoins et les concepts décrits ci-dessus ont été ajoutées. En même temps, certaines fonctionnalités jugées non utiles pour notre cas d'étude ont été enlevées pour ne pas complexifier l'usage de l'outil. Cet outil a été utilisé pour les campagnes opportunistes et flash. Les bases de données sont modifiées en temps réel, et les nouvelles mises à jour sont à la disposition des utilisateurs en temps quasi-réel.

PAYSAGES mobile (téléchargeable depuis Apple et Google Store) est une application mobile qui s'appuie sur les briques logicielles de l'application FotoQuestGo (Laso Bayas et al., 2016). De nouveaux écrans et fonctionnalités ont été implémentés pour répondre aux besoins du protocole décrit ci-dessus. L'application visualise les lieux à visiter et les différents flux WMS. Le contributeur sélectionne un lieu, ensuite il doit se diriger au plus près de ce lieu. Arrivé dans la zone de contribution autorisée, le contributeur peut commencer sa contribution en répondant aux différentes questions posées par l'application, prendre des photos en suivant le guidage de l'application et soumettre les contributions. A tout moment, les contributeurs ont accès à un tableau de bord, décrivant le nombre total de lieux visités, le nombre restant et le classement des contributeurs. Une API a été développée par ma collègue Marie-Dominique Van Damme (LaSTIG) pour permettre à l'application mobile de dialoguer avec le serveur de la plateforme PAYSAGES.

PAYSAGES wiki, développé à l'aide du logiciel Mediawiki, représente la composante sociale de la plateforme PAYSAGES. Il permet à chacun de collaborer facilement en écrivant des informations non structurées. Le wiki a été élaboré grâce au stage de Gabriel Noiret réalisé en 2018 et co-encadré avec mes collègues Laurence Jolivet et Marie-Dominique Van Damme (LaSTIG). Le site wiki PAYSAGES propose une présentation du projet, des guides d'utilisation, une foire aux questions et les dates importantes des campagnes. L'un des principaux défis concernant les données d'occupation de sol est de comprendre les métadonnées. À cette fin, le wiki représente un bon moyen d'aider les contributeurs à comprendre ces informations. Tout d'abord, grâce à l'infobox, des informations sur les types d'usage ou de couverture et des photos illustrant ces types sont directement accessibles depuis le wiki. Tout contributeur peut ajouter une image ou un conseil, ou lancer une discussion dans une page de métadonnées spécifique. La majorité des éléments d'une page de métadonnées sont enregistrés sous forme de propriétés sémantiques. Il convient de noter que le wiki a été utilisé comme preuve de concept pour les campagnes, il n'y a pas eu de maintenance après les campagnes.

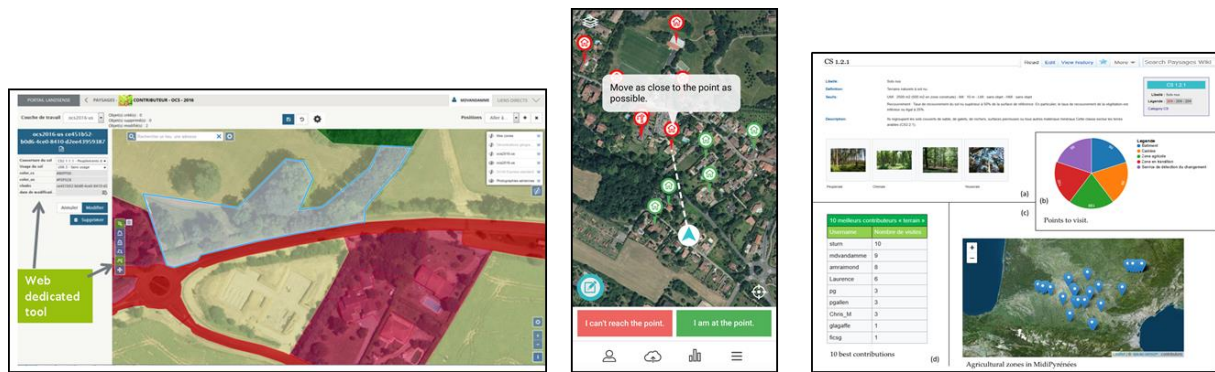
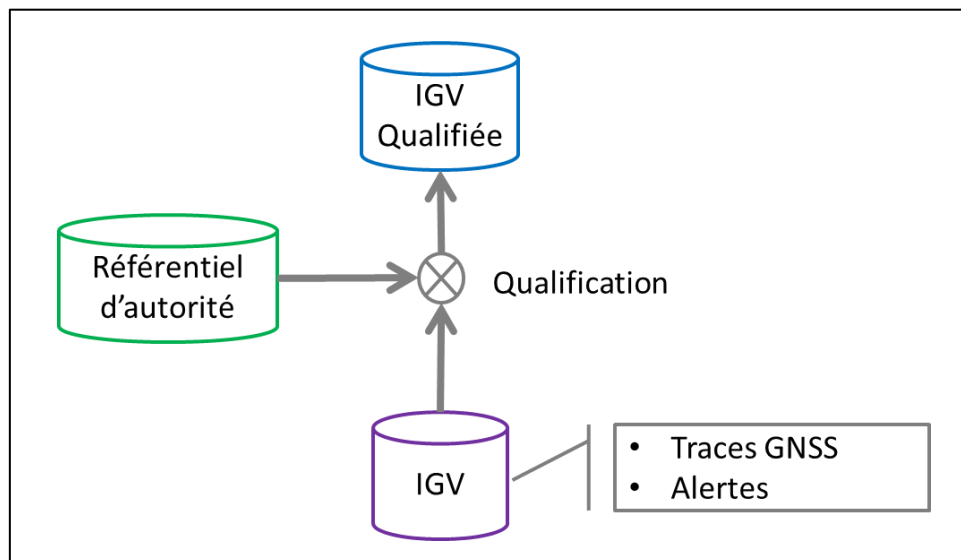


Figure 10. Illustration des outils PAYSAGES : web (à gauche, développé par L. Favral et J. Marcuzzi de l'IGN), mobile (au milieu, développé par T. Sturn de IIASA et M.-D. Van Damme de LaSTIG) et wiki (à droite, développé par G. Noiret et M.-D. Van Damme)

Au total, nous avons organisé dix campagnes dont une campagne guidée (*in-situ*), deux campagnes flash (*in-situ*), 1 campagne opportuniste (bureautique), six campagnes flash (bureautiques). Ces campagnes ont permis l'implication de 141 contributeurs provenant de plusieurs communautés et la collecte de 6937 contributions et 463 photos. Ces différentes informations géographiques volontaires seront décrites plus en détail en Section 4.3.2.

Chapitre 3. Qualification de l'information géographique volontaire



Partie de la démarche globale (Figure 1) décrite dans ce chapitre

Ce troisième chapitre décrit les travaux de recherche qui permettent de répondre au troisième objectif scientifique énoncé dans le Chapitre 1, à savoir **(3) Qualifier et améliorer l'information géographique volontaire** pour faciliter son intégration. Dans le contexte de mes travaux, cet objectif est appliqué à trois types d'IGV : **des alertes opportunistes** issues du système collaboratif de l'IGN, des traces GNSS issues des activités sportives et **des alertes guidées** générées automatiquement via le service de détection de changement ou manuellement et qui ont été validées via un mécanisme collaboratif.

3.1 Problématiques et positionnement

La qualité de l'IGV est une des problématiques qui suscite beaucoup de questions dès lors que l'on s'interroge sur l'usage de l'IGV. De nombreuses publications faisant un état de l'art autour de la qualité de l'IGV ont été publiées (Antoniou et Skopeliti, 2015; Bordogna et al., 2016; Fonte et al., 2015; Fonte et al., 2015; Senaratne et al., 2017; Yan et al., 2020).

La qualité de l'information géographique est un sujet de recherche au cœur de la géomatique depuis des décennies, les premiers travaux sur la qualité ayant été publiés dès les années 1950 (Devillers et al., 2010). Le processus d'évaluation de la qualité de l'IG est de nos jours normalisé par la norme ISO 19157-2013¹⁴. Devillers et Jeansoulin (2005) distinguent deux types de qualité, à savoir la qualité interne (*i.e.* l'adéquation de la base de données géographiques à ses spécifications de contenu) et la qualité externe (*i.e.* l'adéquation de la base de données géographiques pour des usages spécifiques). Six éléments de qualité, nommés aussi indicateurs ou paramètres dans la littérature, sont définis dans la norme ISO 19157-2013 : (1) l'exactitude de position, (2) l'exactitude thématique, (3) la cohérence logique des objets géographiques les uns par rapport aux autres, (4) l'exhaustivité, (5) l'actualité et (6) l'usabilité. Il convient de noter que certains auteurs préfèrent, à juste titre, ajouter l'exactitude sémantique qui se réfère à la nature des objets géographiques afin d'éviter la confusion avec l'exactitude thématique faisant référence aux attributs des objets géographiques. Nous pouvons ajouter à ces indicateurs un autre indicateur qui est souvent utilisé en généralisation et qui porte sur la résolution géométrique : c'est-à-dire la distance minimum entre deux points de la géométrie des objets géographiques (Ruas, 2004).

Premièrement, nous distinguons **les approches extrinsèques d'évaluation de la qualité**. Il s'agit généralement de comparer l'IGV (provenant d'OSM pour la plupart des cas) avec l'information géographique d'autorité ou commerciale. La plupart des travaux suivent la même démarche qui consiste à apparier un jeu de données OSM avec un jeu de données d'autorité, puis de calculer les indicateurs de la norme ISO 19157-2013. Les liens d'appariements de type $n : m$ (n objets d'une base de données correspondent à m objets d'une autre base de données) sont exploités pour calculer les indicateurs d'exactitude géométrique, thématique ou sémantique, les liens de cardinalité $1 : 0$ et $0 : 1$ permettent d'estimer l'exhaustivité géométrique. Les matrices de confusion sont également utilisées pour estimer les exactitudes permettant de calculer les taux d'omission et de commission. En suivant cette démarche, les travaux de recherche évaluent l'exactitude géométrique, thématique et l'exhaustivité (Fan et al., 2014; Girres et Touya, 2010; Haklay, 2010; Jokar Arsanjani et al., 2015a b; Neis et al., 2011; Touya et al., 2017), ainsi que la temporalité et l'évolution (Antoniou et al., 2016). Les approches pour mesurer la cohérence logique reposent généralement sur des règles d'intégrité topologique au sein du même thème (*e.g.* connexité dans les réseaux routiers, partage de géométries entre les frontières voisines) ou inter-thème (*e.g.* une route ne doit pas traverser un bâtiment). Citons, à titre d'exemple, les approches d'estimation de la cohérence logique proposées par (Hashemi et Abbaspour, 2015; Touya et Brando-Escobar, 2013).

Comme le souligne, à juste titre, (Devillers et al., 2019), l'objectif de la qualité interne est « de comparer une représentation parfaite qui représenterait le monde à travers la lentille du modèle à la représentation réellement produite ». Par conséquent, les indicateurs extrinsèques dépendent fortement des modèles des bases de données géographiques comparées. Ceci représente une limite de ces approches dans l'évaluation de la qualité de l'IGV car les différentes sources d'IGV ont des modèles et des objectifs différents.

Deuxièmement, nous distinguons **les approches intrinsèques**. Au sujet du terme *intrinsèque*, il existe un flou terminologique au sein des travaux. D'une part, le terme intrinsèque est utilisé pour définir une caractéristique de l'information géographique. Les approches dites intrinsèques sont basées

¹⁴ <https://www.iso.org/fr/standard/32575.html>

uniquement sur les caractéristiques intrinsèques des objets géographiques de la base de données à évaluer. Les approches intrinsèques proposées s'appuient par exemple sur les métadonnées et l'historique des données (Barron et al., 2014; Mullen et al., 2015). Keßler et de Groot, (2013) proposent, par exemple, de calculer un indicateur de confiance composé de cinq mesures (e.g. le nombre de contributeurs ayant modifié l'objet, le nombre d'éditions par objet). D'autre part, Wang et Strong (1996) emploient le terme *intrinsèque* pour identifier la qualité intrinsèque, qui est un des quatre éléments de la qualité externe impliquant la crédibilité, la précision, l'objectivité et la réputation ; ces mesures de qualité portent donc sur la source d'information dans sa globalité. Dans les travaux d'évaluation de l'IGV, le contributeur est vu comme une source, et ces concepts sont donc transposés pour évaluer la qualité des contributeurs. Ce type d'approche connue sous le nom d'**approche sociale** a été proposé par Goodchild et Li (2012) et s'appuie sur l'hypothèse selon laquelle un ensemble d'objets géographiques produit par un contributeur fiable peut être considéré d'une qualité constante et s'améliorant (Yang et al., 2016). Plusieurs travaux s'intéressent donc à évaluer la qualité du contributeur tels que (Bégin et al., 2018; Bishr et Kuhn, 2007; Flanagan et Metzger, 2008; Truong et al., 2019), pour ne citer que ceux-ci.

Des **approches mixtes (extrinsèque-intrinsèque)** combinant les approches extrinsèques et intrinsèques ont été également proposées. Il s'agit ici de calculer des indicateurs extrinsèques, puis d'utiliser des techniques d'apprentissage pour dériver des indicateurs intrinsèques (Mohammadi et Malek, 2015; Xu et al., 2017).

Enfin, nous pouvons identifier les **approches basées sur le contexte géographique**. Ce type d'évaluation a été initialement proposé par Goodchild et Li (2012) et peut être basé soit sur des lois géographiques telles que la première loi de la géographie de Tobler (Tobler, 1970), soit en croisant l'IGV avec d'autres sources d'information géographique qui peut être volontaire ou d'autorité (Touya et al., 2017; Touya et Brando-Escobar, 2013).

L'IGV que j'ai abordée est hétérogène et variée, obtenue dans des contextes différents et pour des objectifs différents. Ainsi, plusieurs besoins sont identifiés.

Besoins d'évaluer la qualité de l'IGV en l'absence d'IG d'autorité

Un de nos objectifs opérationnels est d'utiliser l'IGV pour la mise à jour de référentiels d'autorité. Pour répondre à cet objectif, j'ai fait le choix de ne pas utiliser l'IGV provenant de la plateforme OSM pour trois raisons principales : (1) éviter les problématiques liées à l'incompatibilité des licences et à la transformation de données suite à des imports massifs dans OSM ; (2) explorer d'autres sources d'IGV telles que les traces issues des sites web ; (3) mieux s'intégrer dans la démarche collaborative de l'IGN et répondre aux besoins d'évaluation de la qualité de l'IGN. Nous avons donc exploré l'usage de trois sources d'IGV : deux sources de type alertes de mises à jour et une source représentant les traces GNSS issues d'activités sportives.

Un premier constat est que les approches d'évaluation de la qualité dépendent du type d'IGV, de ses caractéristiques et du modèle de la base de données qui stocke l'IGV, ainsi que de l'objectif final pour lequel l'usage de l'IGV est souhaité. Dans notre cas de figure, les approches extrinsèques ne sont pas toujours adaptées et ne répondent pas à nos besoins d'estimation de la qualité. En effet, d'une part, il n'est pas souhaitable d'estimer la qualité des traces GNSS en les comparant avec un référentiel d'autorité puisque l'objectif est d'utiliser les traces pour identifier des mises à jour. D'autre part, pour l'information de type alerte, il n'y a pas d'IG volontaire ou d'autorité qui peut servir de référence.

Pour évaluer la qualité des alertes, **une première piste de recherche** est d'utiliser le contexte géographique (Goodchild et Li, 2012) défini par d'autres sources d'IG d'autorité ou volontaire ainsi que les caractéristiques des alertes pour évaluer la qualité de l'alerte. Cette dernière pourrait être exprimée sous la forme d'une confiance telle que proposée par Keßler et De Groot (2013). Cette piste de recherche est explorée en Section 3.2.1. En absence de référence, **l'hypothèse de**

recherche consisterait à dire que les processus collaboratifs permettent, dans un contexte maîtrisé, de constituer des données de référence. Pour y parvenir, notre démarche a été de mettre en œuvre des campagnes de validation des alertes qui sont soumises à plusieurs profils de contributeurs afin d’obtenir des observations multiples ; ensuite ces dernières sont utilisées pour dériver un indicateur de qualité. Cette hypothèse de recherche est étudiée en s’appuyant sur les travaux de Olofsson et al. (2014), Stehman et Foody (2019) et en utilisant des approches basées sur l’estimation d’exactitude sémantique (Stehman et Foody, 2019) et sur l’accord des contributeurs (Fleiss, 1971). Cette hypothèse de recherche est explorée en Section 3.2.2.

Besoin d’évaluer la qualité de l’IGV avec peu de redondances, peu de métadonnées et sans historique d’édition

Un deuxième constat est que la plupart des approches intrinsèques s’appuient sur l’historicité de l’édition des objets géographiques et sur la fiabilité des contributeurs. Bien que transposable dans certains cas, la plupart d’entre elles sont adaptées au modèle de données spécifique de la base de données OSM et à l’édition multiple. L’évaluation de la qualité en se basant sur les approches intrinsèques existantes est donc limitée lorsque la quantité de l’IGV et les caractéristiques intrinsèques provenant de l’historique et des contributeurs sont insuffisantes ou lorsque les métadonnées sont incomplètes ou manquantes. C’est le cas des traces GNSS provenant des activités sportives. En effet, ces traces mises à disposition par le biais des sites web ou des réseaux sociaux sont peu décrites : les contributeurs partagent généralement les informations principales (*e.g.* latitude, longitude, altitude et horodatage de chaque point) sans les métadonnées liées au capteur (*e.g.* le nombre de satellites visibles lors de la collecte des données ou le type de capteur GNSS utilisé). De plus, ces traces sont collectées avec des capteurs différents sans suivre un protocole de collecte ; pour la plupart des sites aucune recommandation n’est donnée. Malgré le fait que la précision des capteurs augmente progressivement au fil des ans, les traces présentent des valeurs aberrantes et parfois une exactitude géométrique insuffisante pour des objectifs tels que la mise à jour du thème routier d’un référentiel d’autorité. Enfin, différents facteurs tels que le relief (Lewis et al., 2007), le couvert végétal (DeCesare et al., 2005), les obstacles (Xie et Petovello, 2015) ont une influence sur l’exactitude et la résolution géométrique. Tous ses aspects font que les traces issues d’activités sportives présentent des hétérogénéités spatiales, thématiques et temporelles intra et inter traces (Ivanovic, 2018). Des approches proposent d’estimer la qualité d’une trace en fonction de la proportion des observations aberrantes (*e.g.* mauvaise exactitude géométrique d’un point GPS, vitesse aberrante). Pour détecter des observations aberrantes, certains auteurs s’appuient sur les filtres de Kalman (Eliasson, 2014; Gomez-Gil et al., 2013) ou sur des propriétés de signal (Iovan et al., 2013). D’autres proposent des approches permettant de calculer des profils médians (Etienne, 2011) ou moyens (Gil de la Vega, 2015) d’un ensemble de traces et d’identifier ainsi des points GNSS aberrants, des segments d’une trace ou des traces aberrantes. Ces approches s’appuyant sur une grande quantité de traces disponibles pourraient contrebalancer l’hétérogénéité spatiale. Cependant, contrairement aux zones urbaines ou péri-urbaines où il y a une grande quantité de traces GNSS, peu de traces qui suivent le même chemin existent en zone de montagne, notre zone d’étude.

Pour évaluer et améliorer la qualité des traces, une **première piste de recherche** a été de s’appuyer sur les caractéristiques intrinsèques des traces en utilisant des indicateurs issus de la littérature tels que la résolution géométrique (Ruas, 2002) ainsi que sur les caractéristiques décrivant le contexte spatial des points GNSS composant la trace. Pour cela, nous nous sommes appuyés sur la littérature qui décrit l’influence de l’acquisition GNSS en fonction de différents contextes : relief, couvert végétal, obstacles. Etant donné l’hétérogénéité intra et inter traces, notre **piste de recherche** est de s’appuyer sur des approches d’apprentissage supervisé, appliquées avec succès dans le domaine de la géomatique (Mustière, 2001; Sheeren, 2005), afin d’identifier les indicateurs pertinents et les seuils de décision dans le but final d’identifier les données aberrantes et estimer l’exactitude

géométrique des traces. Ces deux hypothèses de recherche ont été explorées dans la thèse de S. Ivanovic et présentées en Section 3.3.

3.2 Proposition d'une méthode de qualification de l'information géographique volontaire issue de plateformes d'autorité

Avant-propos

Cette section présente les travaux que j'ai réalisés en collaboration avec ma collègue Laurence Jolivet ainsi que d'autres collègues impliqués dans le projet LandSense. Les partenaires privilégiés ont été : IIASA, l'Université de Nottingham et de Heidelberg ainsi que la compagnie Géoville. Deux aspects y sont traités. D'une part, les travaux de qualification de l'IGV collectée via une première initiative proposée par l'IGN à travers le système Ripart. D'autre part, les travaux de qualification de l'IGV collectée via la plateforme PAYSAGES. Le point commun de ces travaux réside dans le fait que l'IGV provient des plateformes mises en place par l'IGN. La première propose un protocole simple et considère l'IGV comme une alerte pour la mise à jour des référentiels d'autorité. La deuxième, quant à elle, s'appuie sur le protocole et les recommandations décrits en Section 2.3 et considère l'IGV comme une source d'information qui peut être intégrée pour mettre à jour le référentiel d'autorité.

Cette section résume donc les travaux des deux publications suivantes:

Jolivet, L., et Olteanu-Raimond, A-M, 2017. 'Crowd and Community Sourced Data Quality Assessment'. In *Advances in Cartography and GIScience*, edited by Michael P. Peterson, 47–60. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57336-6_4.

Olteanu-Raimond, A.-M., L. See, M. Schultz, G. Foody, M. Riffler, T. Gasber, L. Jolivet, et al. 2020. 'Use of Automated Change Detection and VGI Sources for Identifying and Validating Urban Land Use Change'. *Remote Sensing* 12 (7): 1186.

3.2.1 Méthode d'estimation de la confiance des alertes opportunistes

L'IGV dont il est question dans cette section est issue de la plateforme Espace Collaboratif proposée par l'IGN. Il s'agit d'une information qui est collectée de manière opportuniste, c'est-à-dire que le contributeur peut participer à tout moment et que la collecte n'est pas guidée par un objectif ou un lieu précis. La plateforme s'appuie sur un protocole de collecte qui utilise la notion d'alerte. Le contributeur localise l'alerte via un service de localisation et remplit des attributs à travers un formulaire en ligne pour la décrire. Il peut également dessiner un croquis pour montrer l'étendue de l'erreur ou du changement et joindre des documents (*e.g.* photos, documents) pour justifier l'erreur ou le changement remarqué. Enfin, il peut soumettre l'alerte et reçoit un courriel de confirmation.

Plus précisément, chaque alerte est caractérisée par une localisation ponctuelle (latitude, longitude) et un ensemble d'attributs : la commune concernée par l'alerte, le thème et un commentaire. Il convient de noter que le commentaire est un champ obligatoire et ouvert qui laisse aux contributeurs la liberté de décrire plus en détail l'alerte tandis que le nom de la ville et le thème sont normalisés, les valeurs devant être choisies à partir de listes prédéfinies. Le thème s'appuie sur le modèle de données de la BDTOPO® et sur un thème métier qui permet de répondre aux besoins des partenaires publics (*e.g.* les pompes à incendie pour les services d'incendie). Notons également que le protocole de collecte a connu des changements au fil du temps, et que ceux-ci concernent principalement la définition des thèmes (*i.e.* choix multiples, deux choix possibles et enfin un seul choix dans la version actuelle) et le nombre de thèmes présents dans la liste.

Pour pouvoir contribuer, l'inscription est obligatoire, seuls une adresse e-mail valide et un mot de passe sont requis. Le contributeur a la possibilité de participer d'une façon indépendante ou de rejoindre un ou plusieurs groupes. En plus des caractéristiques décrites ci-dessous, les alertes sont

également caractérisées par des métadonnées relatives au contributeur, telles que l'identifiant du contributeur, la date, l'appartenance à un groupe, le type de support utilisé (*e.g.* application web, plug-in), la version du protocole de collecte utilisée, le nombre de croquis et de fichiers joints. Lors de la soumission, chaque alerte a le statut de « soumise » par défaut.

En ce qui concerne le traitement des alertes, celles-ci sont d'abord vérifiées et validées manuellement par les collecteurs (agents de l'IGN) en s'appuyant sur les spécifications de la BDTOPO®, des orthophotos et d'autres sources de données externes telles que les permis de construire délivrés par les municipalités. A travers cette vérification, le collecteur peut changer le statut de l'alerte (*i.e.* soumise, en cours de vérification ou vérifiée). Cette tâche de vérification et de validation est très coûteuse en temps.

Dans ce contexte, nous nous sommes intéressés à la qualification des alertes. Notre objectif a été de proposer une approche permettant de mesurer d'une manière automatique la confiance des alertes. La confiance d'une alerte est définie par un ensemble d'indicateurs qui mesurent la cohérence entre une alerte et son contexte spatial. La cohérence intrinsèque consiste à comparer les différentes caractéristiques d'une même alerte tandis que la cohérence contextuelle s'appuie sur l'approche géographique proposée par Goodchild et Li (2012) et compare les caractéristiques d'une alerte avec son contexte spatial. Dans un contexte opérationnel d'aide à la décision, les indicateurs de cohérence peuvent être combinés pour dériver une confiance globale pour chaque alerte.

Etant donné les caractéristiques des alertes collectées, nous avons proposé cinq indicateurs (voir le Tableau 2) : trois d'entre eux sont contextuels (To&L, Th&L et Co&L), un est intrinsèque (Th&Co), un est mixte (FContributeur). Bien que les indicateurs s'appuient sur des mesures quantitatives, le résultat final est, quant à lui, une mesure qualitative exprimant la cohérence à travers trois valeurs : oui/non/peut-être.

Cohérence To&L. Cet indicateur mesure la cohérence entre l'appartenance de l'alerte à une commune choisie par le contributeur (NomC) et le nom de la commune dans laquelle l'alerte a été localisée (NomL). Ainsi, pour chaque alerte, nous récupérons le nom de la commune d'appartenance (NomC) en fonction de sa localisation (x,y). Ensuite, la comparaison consiste à calculer une mesure de similarité entre les deux noms.

Cohérence Th&L. Cet indicateur mesure la cohérence entre le thème choisi par le contributeur (ThèmeC) et le contexte spatial de l'alerte. Ce dernier s'appuie sur la notion de voisinage et est déterminé à partir de la BDTOPO®. Pour chaque alerte, les objets géographiques voisins représentés dans la BDTOPO® sont sélectionnés grâce aux opérateurs topologiques et métriques. Ainsi, le contexte spatial d'une alerte est défini par l'ensemble des thèmes d'objets sélectionnés :

ThèmeCS = {ThèmeCS₁, ... ThèmeCS_i}, où $i = 1 \dots N$ (N est le nombre d'objets géographiques de la BDTOPO® constituant le contexte spatial de l'alerte).

Ensuite, une distance sémantique est calculée entre le thème choisi par l'utilisateur ThèmeC et chacun des thèmes ThèmeCS_i des objets composant le contexte spatial de l'alerte.

Cohérence Co&L. L'indicateur mesure la cohérence entre le commentaire effectué par le contributeur {CommentaireC} et le contexte spatial de l'alerte. Etant donné que le commentaire est un champ libre, une grande variété de mots peut apparaître. Par conséquent, une première étape a consisté à définir un dictionnaire qui permet d'associer à chaque thème de la BDTOPO® une liste de concepts issus d'une taxonomie des concepts topographiques (Mustière et al., 2011a). Par exemple, les concepts associés au thème Bâti peuvent être maison, mairie, etc. Une deuxième étape a consisté à réaliser un prétraitement de chaque commentaire afin d'en extraire une liste des mots signifiants du commentaire {CommentaireC}_k, $k=1..K$. Le contexte spatial et le calcul des distances est similaire à l'indicateur Cohérence Th&L sauf que le calcul des distances sémantiques s'effectue entre l'ensemble des mots signifiants du commentaire {CommentaireC}_k et l'ensemble des éléments du dictionnaire associés aux thèmes composant le contexte spatial de l'alerte.

Cohérence Th&Co. Cet indicateur intrinsèque mesure pour chaque alerte la cohérence entre le commentaire {CommentaireC} et le thème (ThèmeC) choisis par le contributeur. De la même manière que pour l'indicateur Co&L, le commentaire est divisé en mots signifiants du commentaire {CommentaireC}_j, j=1..M, et un dictionnaire est associé au (ThèmeC). Les distances sémantiques sont calculées entre l'ensemble des mots signifiants du commentaire et l'ensemble des éléments du dictionnaire associés au thème attribué par le commentaire.

Fiabilité contributeur. L'idée de cet indicateur était de mesurer la fiabilité d'un contributeur en se basant sur les indicateurs précédents. L'hypothèse consiste à considérer que la fiabilité d'un contributeur se mesure par le nombre de contributions cohérentes, ainsi que par des caractéristiques sociales telles que l'expertise, l'appartenance au groupe, la longueur de la contribution, comme suggéré par (Foody et Boyd, 2012; Van Exel et al., 2010). Il convient de noter que cet indicateur n'a pas été implémenté. Les règles permettant d'estimer la cohérence d'une alerte sont décrites dans le Tableau 2.

Tableau 2. Définition des règles pour mesurer la confiance des alertes

Indicateurs	Mesures	Règles
Cohérence To&L	Distance de Levensteihn entre chaînes de caractères (<i>dString</i>) distance euclidienne min (<i>dE_{min}</i>) entre la position de l'alerte et la frontière de la commune d'appartenance	1 : Si NomL non nulle et <i>dString</i> ≤ seuil → To&L = Oui 2 : Si NomL non nulle et <i>dString</i> > seuil → To&L = Non 3 : Si NomL est nulle et <i>dE_{min}</i> ≤ seuil → To&L = Peut-être 4 : Si NomL est nulle et <i>dE_{min}</i> > seuil → To&L = Non
Cohérence Th&L	distance sémantique basée sur une taxonomie { <i>dSémantique_i</i> } i=1..N	1 : Si au moins une { <i>dSémantique_i</i> } ≤ seuil, i=1..N → Th&L = Oui 2 : Si toutes { <i>dSémantique_i</i> } > seuil, i=1..N → Th&L = Non 3 : Si pas de contexte spatial déterminé → Th&L = Peut-être
Cohérence Co&L	distance sémantique basée sur une taxonomie { <i>dSémantique_k</i> }, k=1..K, où K est le nombre de distances sémantiques calculées	1 : Si au moins une { <i>dSémantique_k</i> } ≤ seuil, k=1..K → Co&L = Oui 2 : Si toutes { <i>dSémantique_k</i> } > seuil, k=1..K → Co&L = Non 3 : Si contexte spatial non-déterminé OU non existence des mots significatifs dans le commentaire → Co&L = Peut-être
Cohérence Th&Co	distance sémantique basée sur une taxonomie { <i>dSémantique_j</i> }, j=1..M, où M est le nombre de distances sémantiques calculées	1 : Si au moins une { <i>dSémantique_j</i> } ≤ seuil, j=1..M → Th&Co = Oui 2 : Si toutes { <i>dSémantique_j</i> } > seuil, j=1..M → Co&L = Non 3 : Si non existence des mots significatifs dans le commentaire → Th&Co = Peut-être

L'approche de l'estimation de la confiance est illustrée en Figure 11. Chaque alerte est qualifiée par les indicateurs décrits ci-dessus à travers une cohérence qui peut prendre trois valeurs qualitatives : oui, peut-être, et non. Chaque alerte qualifiée (QReport) peut être décrite par chaque indicateur séparément ou par une combinaison de ceux-ci. La combinaison est déterminée en prenant en compte le nombre de qualificatifs obtenus pour chaque indicateur et qualifiera chaque alerte selon trois valeurs : cohérente, moyennement cohérente, et incohérente.

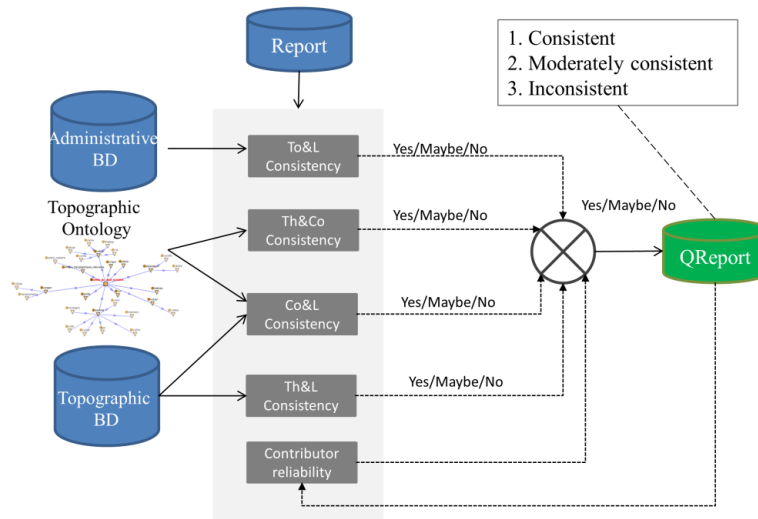


Figure 11. Approche pour l'estimation de la confiance d'une alerte ; issue de (Jolivet et Olteanu-Raimond, 2017)

3.2.1.1 Application et résultats

L'approche décrite ci-dessus a été appliquée sur 59680 alertes collectées via la plateforme *Espacecollaboratif* (ancienne Ripart) de l'IGN entre 2009 et 2015. Ces alertes concernent 3766 communes représentant environ 10% du territoire français de métropole et d'outre-mer. La répartition spatiale des alertes est très inégale entre les communes ainsi qu'entre les départements. Par exemple, 20% des départements ont plus de 100 alertes et seulement 5% ont plus de 1000 alertes. Cette hétérogénéité spatiale est directement liée aux groupes de contributeurs et à l'intérêt qu'ils portent à la plateforme collaborative.

Concernant la variabilité des types d'alerte (le thème choisi par les contributeurs), nous avons observé 32 types distincts. Parmi ceux-ci, 76 % des alertes sont associées à des types *métier* définis pour les besoins des groupes de contributeurs (e.g. le type *DECI* utilisé par les pompiers pour remonter des alertes sur les pompes à incendie). La Figure 12 illustre la fréquence des thèmes associés aux alertes, excepté ceux qui sont dédiés aux besoins professionnels. Les thèmes les plus fréquents sont les adresses ou les lieux, les routes, les bâtiments et les POI (Points d'Intérêt). Il est important de noter que certaines alertes n'ont pas de thème spécifique associé (e.g. Bases de données IGN, Autre, ou Null).

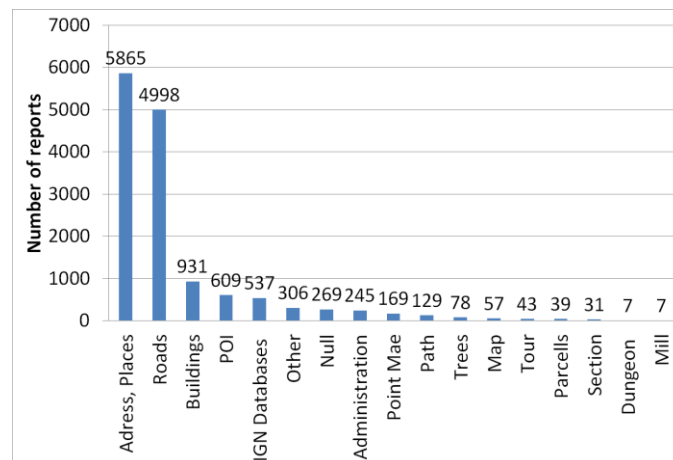


Figure 12. Fréquence des thèmes associés aux alertes pour la période 2009-2015 ; issue de (Jolivet et Olteanu-Raimond, 2017)

Enfin, concernant les commentaires, seulement 26% des alertes en possèdent un. Les mots les plus fréquents concernent des éléments administratifs et des objets topographiques (*e.g.* ville, rue, ruelle, hôtel de ville), des relations spatiales (en direction de, devant), des toponymes et sont parfois liés à la collecte de données (*e.g.* source, nouveau, modification). L'analyse conjointe des thèmes et des commentaires a mis en évidence le fait que les deux caractéristiques sont riches en information et parfois complémentaires (*e.g.* pour une alerte dont le thème est *Autre* ou *Null*, le commentaire permet de mieux caractériser l'alerte). Quelques résultats seront présentés par la suite.

La Figure 13 illustre un exemple de la cohérence des alertes (Yes, Maybe et No) pour la commune de Concarneau en utilisant l'indicateur de cohérence To&L. Parmi les 48 alertes, 41 alertes sont cohérentes, 6 non cohérentes et une moyennement cohérente. Cette dernière est due au fait qu'elle correspond à des éléments du paysage en dehors des limites administratives. Les alertes non cohérentes sont pour la plupart proches de la frontière de la commune cible. Cela pourrait s'expliquer soit par une erreur de position de l'alerte, soit par une erreur d'attribution de la commune d'appartenance. Le changement des limites communales ne nous semblait pas être une hypothèse dans notre cas.

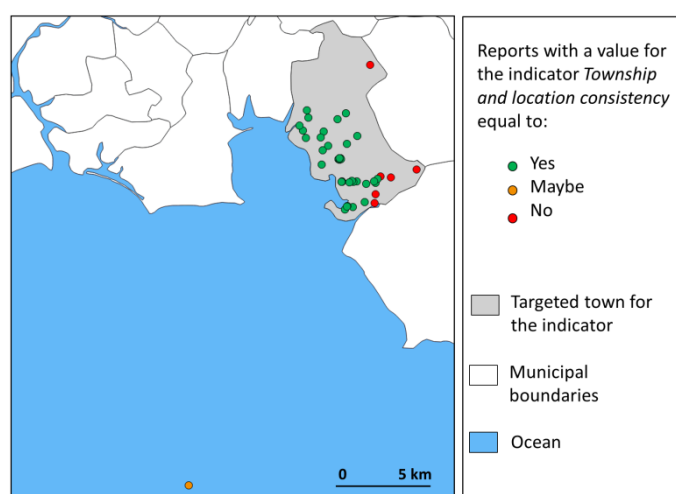


Figure 13. Cohérence des alertes de la commune de Concarneau selon l'indicateur de cohérence To&L (appartenance à une commune choisie par le contributeur et localisation ponctuelle de l'alerte)

La Figure 14 illustre, quant à elle, trois résultats obtenus selon l'indicateur Th&L (cohérence entre le thème de l'alerte et les thèmes de son contexte spatial).

L'exemple de la Figure 14b) permet surtout d'identifier une alerte qui n'est pas compatible avec la BDTOPO car elle ne contient pas les pompes d'incendies.

Les résultats ont été validés manuellement pour le département Pas-de-Calais qui comptabilise à lui seul 75 % des alertes. D'une manière générale, nous avons pu remarquer que la majorité des alertes sont considérées comme cohérentes selon l'indicateur To&L. Les alertes non cohérentes correspondent à des éléments de paysage situés en dehors de toute limite communale (*e.g.* nouveau pont au-dessus d'un estuaire, plage, ligne de flottaison dans une enclave étrangère). L'analyse des résultats produits par cet indicateur nous mène à penser qu'il n'est pas discriminant pour l'évaluation de la cohérence. En revanche, des alertes identifiées comme non cohérentes sont utiles pour détecter des mises à jour (*i.e.* modification du paysage ou nouvelle construction).

L'usage du champ commentaire s'est avéré être prometteur et discriminant : environ 50 % des alertes possèdent un commentaire contenant des mots significatifs. L'indicateur intrinsèque Th&Co (cohérence thème et commentaire) a permis d'identifier 6% d'alertes non cohérentes. En revanche, la moitié des alertes présentent un commentaire qui n'a aucune signification. En effet, étant donné que dans le protocole de collecte le champ commentaire est obligatoire, de nombreux contributeurs le remplissent avec des lettres n'ayant aucune signification (*e.g.* ajksjisfi, aaaa, klmk) Ceci rend notre

processus de prétraitement de texte très difficile et peut parfois introduire un biais dans le résultat final. Notre recommandation est de rendre le champ commentaire optionnel. Ainsi, les contributeurs n'ayant pas la motivation de donner des précisions ne seront pas obligés de le remplir.

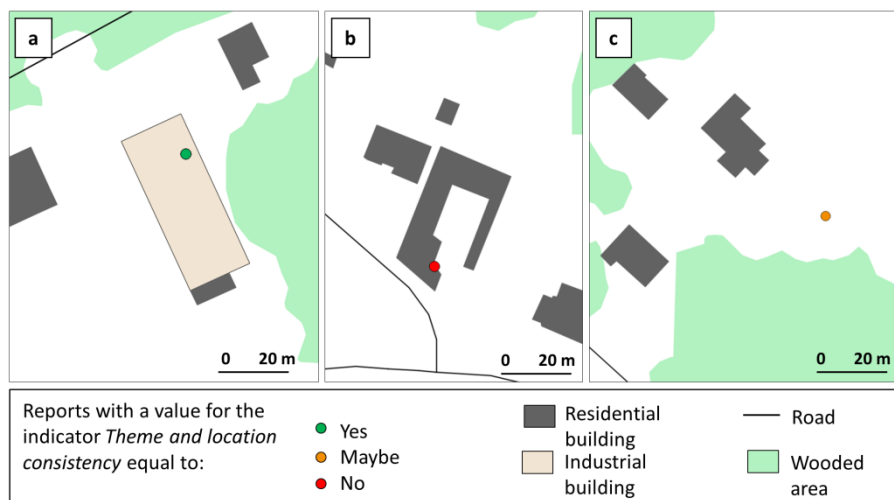


Figure 14. Cohérence des alertes pour l'indicateur Th&L (Thème de l'alerte et les thèmes du contexte spatial de l'alerte) : (a) Thème de l'alerte (Bâti) cohérent avec les thèmes de son contexte spatial (Bâti) ; (b) Thème de l'alerte (pompe d'incendie) non cohérent avec les thèmes de son contexte spatial (Bâti) ; (c) Alerte sans contexte spatial déterminé, donc cohérence moyenne.

Le travail présenté dans cette section représente une première étape dans la qualification de l'information géographique volontaire de type alerte. D'une part, ce travail nous a permis de mieux connaître les contributeurs de notre plateforme, la manière dont ils contribuent en groupe, leur motivation et ainsi de dégager des recommandations pour le protocole de collecte de l'IGV (voir Section 2.3). D'autre part, nous avons pu identifier les faiblesses de cette approche qualitative et dégager de nombreuses pistes d'amélioration telles que la modélisation des connaissances en prenant mieux en compte l'imprécision et l'incomplétude, la combinaison des indicateurs qui pourrait être basée sur des méthodes de fusion permettant un passage plus précis des indicateurs quantitatifs aux indicateurs qualitatifs. Concernant les indicateurs eux-mêmes, le contexte spatial pourrait être amélioré en s'appuyant sur d'autres sources de données (*e.g.* occupation du sol, photos prises sur le terrain) ou en prenant en compte non seulement des opérateurs topologiques, mais aussi des distances. Les indicateurs basés sur l'analyse des commentaires, très ad-hoc dans cette première approche, pourraient également être améliorés en s'appuyant sur des outils de traitement de langage. Enfin, nous pouvons aller plus loin et assister l'opérateur dans sa prise de décision finale. Les alertes identifiées ayant une confiance forte ou moyenne pourraient être validées automatiquement et par conséquent le statut de l'alerte pourrait être changé en « approuvée » ou « refusée », tandis que l'alerte identifiée comme ayant une faible confiance pourrait générer une vérification manuelle à l'opérateur. Par conséquent, la combinaison des indicateurs doit être adaptée aux besoins des opérateurs en étant plus ou moins permissive.

3.2.1.2 Analyses des contributeurs

Les contributeurs de la plateforme *Espacecollaboratif* pour la période 2009-2015 sont majoritairement des agents appartenant à des organismes publiques telles que le service départemental d'incendie et de secours (SDIS), les mairies, les collectivités locales. Parmi les 520 contributeurs, 46% appartiennent à un groupe (*e.g.* SDIS 62, SDIS 78) et participent dans le cadre de leurs activités professionnelles (bureautique ou sur le terrain) ou pendant leur temps libre. D'autres groupes identifiés dans notre jeu de données correspondent à des actions ciblées : initiative entre plusieurs organismes et campagnes à l'intention des citoyens pendant une courte période. Le

nombre de contributeurs par groupe est hétérogène, variant de 1 à 277 contributeurs (moyenne de 13 et écart-type de 45).

Concernant les contributeurs et leur motivation, la répartition présentée en Figure 15 montre que 20% des contributeurs ont réalisé environ 90% des alertes avec plus de 1000 alertes par contributeur. Ce résultat montre que, même si elle est relativement comparable, la participation des contributeurs ne correspond pas au comportement connu rencontré dans d'autres projets collaboratifs tels que Wikipédia, la « loi » de « 90-9-1 ». Dans la communauté OSM, cette « loi » est de « 81-14-5 » (Neis et Zipf, 2012). Dans notre cas, nous estimons une « loi » de « 65-15-20 ».

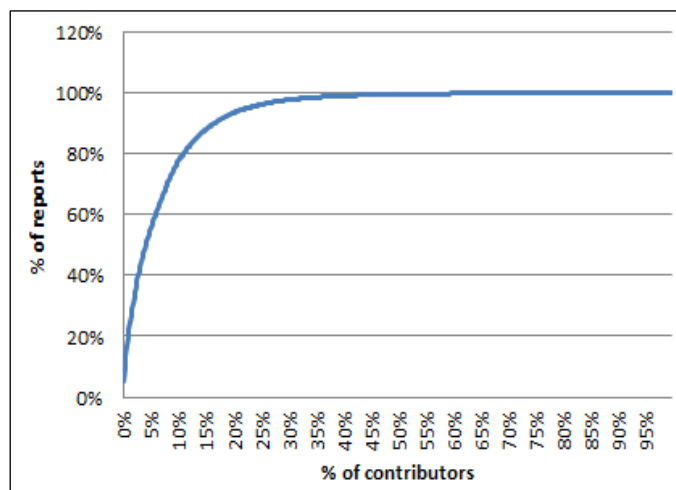


Figure 15. Répartition du pourcentage de contributions par contributeur ; issue de (Jolivet et Olteanu-Raimond, 2017)

En ce qui concerne l'activité des groupes, nous avons constaté un groupe très actif qui a effectué 77% de la totalité des alertes. À l'intérieur du groupe, le pourcentage de contributeurs inactifs suit la même loi « 65-15-20 ».

Nous avons également analysé la relation entre le nombre de contributeurs par groupe et le nombre d'alertes effectuées par le groupe d'une part, et le nombre d'alertes et la durée, d'autre part. Nous avons constaté une faible corrélation positive entre le nombre de contributeurs par groupe et le nombre d'alertes ($r^2=0,13$, $p\text{-value}<0,05$). Cependant, si nous enlevons de l'analyse les deux groupes qui fournissent la plupart des alertes, il n'y a plus de corrélation. Quant à la relation entre la durée de participation et le nombre d'alertes, il n'y a pas de relation étroite, mais il existe une faible corrélation positive significative (valeur de Pearson = 0,37, intervalle de confiance de 99,9 %). Précisons que la durée de participation par contributeur correspond à l'intervalle de temps entre le premier jour de contribution et le dernier jour et varie entre 1 et 2500 jours. Au niveau des groupes, cinq groupes n'ont contribué que pendant une journée, 23 groupes pendant plus d'un an. La plupart des groupes (6) ont une continuité depuis le début de l'année 2013. Parmi tous les groupes, 31 d'entre eux contribuent encore pendant la dernière année de l'étude. Le groupe le plus actif est le groupe SDIS 62 du département Pas-de-Calais qui a commencé sa contribution en 2013 et compte à ce jour 125 contributeurs (23%).

3.2.2 Méthode d'estimation de la qualité des alertes générées automatiquement

Cette section présente les travaux, publiés dans (Olteanu-Raimond et al., 2020), que j'ai effectués en collaboration avec des collègues impliqués dans le projet LandSense.

Un Service de Détection de Changements (SDC) d'usage du sol en zone urbaine et périurbaine a été développé dans le cadre du projet LandSense. Le service est basé sur une méthode de détection de changement qui analyse des séries temporelles provenant des images Sentinel 2 et qui fournit en sortie des changements représentant les nouvelles constructions. Les changements sont représentés

sous forme de polygones et sont classifiés en quatre classes d'usage : Résidentiel, Industriel, Infrastructure et Autre. Nous considérons les changements comme des alertes qui ont pour vocation d'être utilisées pour la mise à jour de la base de données d'occupation du sol, OCS-GE® produite par l'IGN.

Dans ce contexte, notre objectif a été de proposer une méthodologie pour valider et qualifier les changements détectés. Pour plus d'information sur la méthode de détection de changements, que je ne décris pas ici, le lecteur est invité à consulter la publication (Olteanu-Raimond et al., 2020 a).

3.2.2.1 Méthode de validation collaborative des changements d'usage de sol

La méthodologie de validation proposée s'inspire des recommandations proposées par les travaux de la communauté de télédétection pour estimer l'exactitude sémantique de l'occupation des sols (Olofsson et al., 2014; Stehman et Foody, 2019) et implémente le protocole décrit en Section 2.3.2. La méthode de validation est composée de trois étapes principales : l'échantillonnage, la conception de la réponse et la validation elle-même.

Choix de l'échantillonnage. Cette étape consiste à sélectionner les alertes à valider. En utilisant l'imagerie Sentinelle 2 pour la période 2015-2019, le service de détection des changements a étudié une zone d'environ 10 000 km² représentant trois départements français : le Gers, les Hautes-Pyrénées et la Haute-Garonne (voir Figure 13). Etant donné l'objectif final, qui est de mettre à jour l'OCS-GE, nous avons eu recours à un échantillonnage spatio-temporel. Ainsi, d'une part, nous avons sélectionné les alertes dans la zone du SCOT Toulousain (Schémas de Cohérence Territoriale de la grande agglomération de Toulouse) qui fera l'objet d'un nouveau millésime 2019 produit par l'IGN et publié en 2020. D'autre part, seules les alertes représentant des changements apparus dans la période 2016-2019 sont sélectionnées afin de quantifier les changements entre les deux millésimes 2016 et 2019.

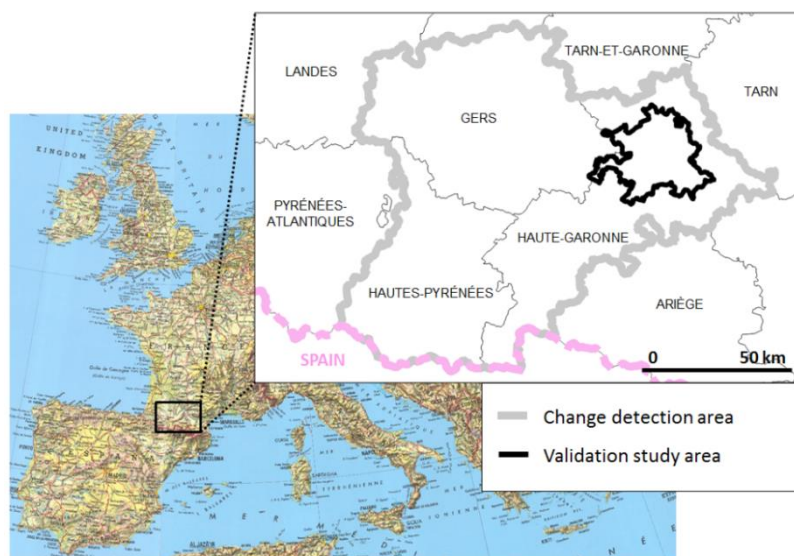


Figure 16. Limites des zones d'étude pour la détection de changements et la validation (Fond cartographique: GEOFLA Geoportal©IGN) ; issue de (Olteanu-Raimond et al., 2020 a)

Conception de la réponse. Cette deuxième étape consiste à définir le protocole qui permettra la validation des alertes d'une manière collaborative. Plus précisément, il définit la manière dont les contributeurs attribueront une classe à chaque changement détecté, le type de données supplémentaires à utiliser dans les outils afin de faciliter l'interprétation, et la liste d'étiquettes supplémentaires à ajouter pour réduire la confusion entre les différentes classes d'usage pendant le processus de validation.

Pour éviter d'influencer la décision du contributeur, nous avons choisi une validation à l'aveugle, c'est-à-dire que la question et les données en entrée sont préparées de manière à ce que les contributeurs ne connaissent pas la classe d'usage attribuée automatiquement par le SDC.

Comme mentionné ci-dessus, les changements sont classés par le SDC en quatre classes d'usage : Résidentiel, Industriel, Infrastructure et Autre. Étant donné que certaines nouvelles constructions ont lieu sur une longue période et que leur usage ne peut être identifié qu'à la fin des travaux, nous avons ajouté l'étiquette *Construction en cours* et *Destruction*. Afin de prendre en compte les sur-détections (i.e. changement détecté à tort), nous avons ajouté une troisième étiquette *Pas de changement*. Enfin, compte tenu des difficultés à valider un changement ou sa classe d'usage, l'étiquette *Inconnu* a été ajoutée. Cela permet aussi aux contributeurs de ne pas se prononcer sur un changement, au lieu de leur imposer un choix.

Concernant les informations supplémentaires permettant aux contributeurs de choisir une étiquette appropriée à chaque changement, plusieurs flux WMS ont été ajoutés dans l'outil de validation. Ainsi, la validation a été réalisée par photo-interprétation en comparant deux orthophotos de 2016 et 2019 de résolution comprise entre 20 et 50 cm. De plus, nous avons intégré dans les outils de validation d'autres flux WMS tels que le Scan 25, les dénominations géographiques, etc.

Définition d'une stratégie pour l'organisation des campagnes de validation. Pour ce travail de validation collaborative, nous avons implémenté le concept de *campagne flash* (Section 2.3.2), c'est à dire une campagne de courte durée comprise entre 2 et 4 heures. Chaque campagne flash s'est déroulée en utilisant l'outil bureautique Paysages décrit en Section 2.3.2 et dans les mêmes conditions : les mêmes alertes et flux WMS en entrée, les mêmes objectifs, et des explications d'aide à la validation en début de chaque campagne. Pour chaque campagne, un à deux membres de l'équipe projet de l'IGN ont fourni une assistance. Nous avons organisé sept *campagnes flash*, chacune impliquant un groupe de contributeurs ayant des profils différents :

- Profil 1 : Etudiants de 1^{ère} année de l'École Nationale des Sciences Géographiques (ENSG) qui avaient suivi une formation sur la photo-interprétation avant la *campagne flash*. L'exercice de validation a été inclus dans leur programme d'études afin d'appliquer les compétences liées à la photo-interprétation. Ce groupe composé de 60 étudiants a été divisé en deux groupes, comprenant chacun 30 étudiants et identifiés plus loin comme Profil 1-gr1 et Profil 1-gr 2.
- Profil 2 : Etudiants du Master 2 en Géomatique (M2IGAST) de l'ESNG-Université Gustave Eiffel. La *campagne flash* a été également inscrite dans leur programme d'études et les alertes ont été utilisées par la suite lors d'une séance de travaux pratiques. Ces étudiants ont une formation en télédétection, classification spatiale et validation des données géographiques dans le cadre de leur cursus, avant la *campagne flash*. Parmi les 19 étudiants, certains avaient également une expérience en photo-interprétation.
- Profil 3 : Étudiants de 1^{ère} année de l'ENSG n'ayant pas reçu de formation en photo-interprétation. Ce groupe était composé d'un ensemble de 15 étudiants volontaires, qui venaient de commencer leur cursus en sciences géographiques. Ils étaient motivés par une meilleure compréhension des applications environnementales liées à l'artificialisation des sols, du climat et de la biodiversité et par le lien avec la recherche. L'un des étudiants était originaire de la zone d'étude.
- Profil 4 : Quatre chercheurs en géomatique, experts des données d'occupation du sol du laboratoire LASTIG de l'IGN-France. Les contributeurs de ce groupe connaissaient également le fonctionnement du SDC de LandSense.
- Profil 5 : Experts des données d'occupation du sol provenant des autorités locales. Ce groupe était composé de cinq agents de l'Agence Urbaine d'Aménagement de Toulouse (AuAT) ayant une expertise en photo-interprétation, OCS-GE et saisie des données géographiques, mais une faible expertise en télédétection.

- Profil 6 : Agents administratifs de l'IGN. Ce groupe était composé de deux agents administratifs de l'IGN sans expertise en données géographiques ou photo-interprétation. Au total, 105 contributeurs ont été impliqués et 2778 observations ont été collectées.

3.2.2.2 Qualification des alertes générées automatiquement

Pour qualifier les observations collectées nous avons proposé deux indicateurs de qualité : l'accord des contributeurs et l'exactitude sémantique des classes d'usage.

Accord des contributeurs. Une des richesses de l'IGV collectée est la redondance. En effet, grâce à la stratégie définie pour la mise en place des campagnes, chaque alerte de changement a été validée par au moins trois contributeurs (voir la Figure 17). Par exemple 159 alertes ont été validées par 3 contributeurs et 100 alertes ont été validées par six contributeurs différents.

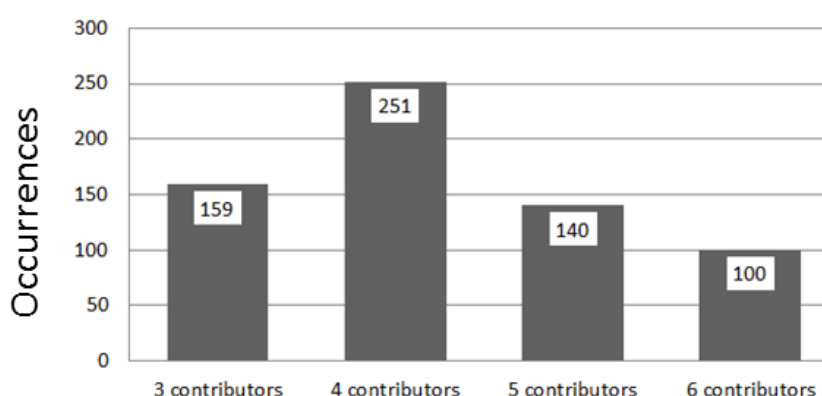


Figure 17. Le nombre total de changements validés en fonction du nombre de contributeurs ; issue de (Olteanu-Raimond et al., 2020)

Afin d'estimer à quel degré les contributeurs ayant différents profils d'expertise sont en accord ou en désaccord, nous avons proposé un indicateur basé sur le coefficient kappa de Fleiss (Fleiss, 1971) et qui prend en compte le nombre de contributeurs par alerte et le nombre d'étiquettes possibles. L'accord des contributeurs est compris entre 0 et 1, 0 signifiant désaccord total et 1 accord total. Dans la Figure 18, nous observons que la grande majorité des cas de désaccord total se produisent lorsque les alertes sont validées par trois contributeurs et que la plus grande partie des cas d'accord très élevé viennent de situations où trois ou quatre contributeurs se prononcent.

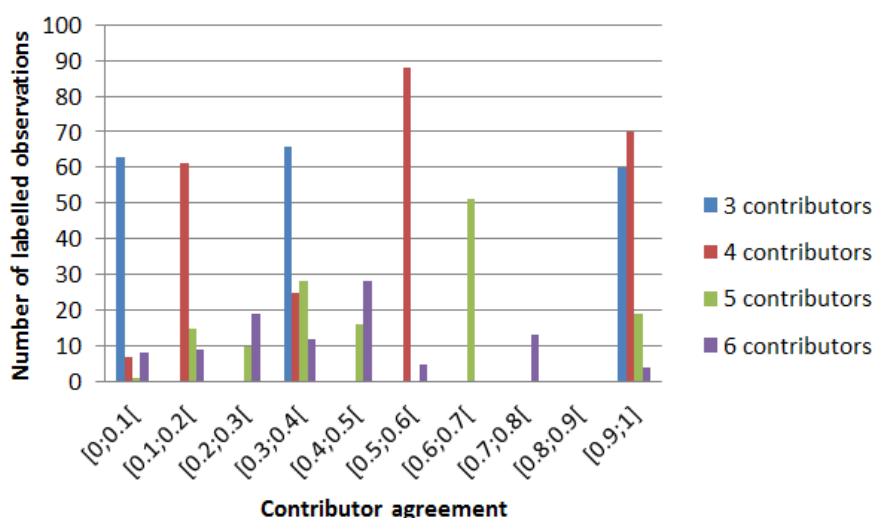


Figure 18. Fréquence du degré d'accord du contributeur pour les alertes validées ; issue de (Olteanu-Raimond et al., 2020)

Exactitude sémantique. L'estimation de l'exactitude sémantique a été calculée grâce à la matrice de confusion. Cette dernière est calculée en comparant pour chaque alerte sa classe d'usage attribuée par le SDC et la classe d'usage de référence. Les classes d'usage de référence ont été calculées à partir des validations multiples. Tout d'abord, nous n'avons sélectionné que les alertes validées dont l'accord des contributeurs est supérieur à un seuil choisi empiriquement (dans ce travail il a été fixé à 0.4). Ensuite, nous avons déterminé pour chaque alerte sélectionnée la classe d'usage de référence. Cette dernière est associée à la classe dominante (*i.e.* choisie par une majorité de contributeurs). En cas d'égalité, la classe d'usage de référence est choisie au hasard.

A partir de la matrice de confusion, nous avons calculé l'exactitude sémantique globale (*i.e.* la proportion des alertes correctement classées par rapport à la classification de référence), l'exactitude utilisateur (*i.e.* la proportion des alertes correctement classées pour chaque classe par rapport au nombre d'alertes classées dans cette classe par le SDC) et l'exactitude producteur (*i.e.* la proportion des alertes correctement classées pour chaque classe par rapport au nombre d'alertes classifiées dans cette classe dans la classification de référence).

L'exactitude globale obtenue est égale à 81%. Nous avons remarqué que les exactitudes utilisateur et producteur sont supérieures à 85% pour la classe Résidentielle. Cela signifie que le SDC est approprié pour la détection des changements résidentiels. En revanche, l'exactitude utilisateur pour les classes Infrastructure et Industrielle est faible, étant égale respectivement à 50% et 70%. Cela n'est pas surprenant, étant donné la difficulté à détecter ces deux types d'usage.

La matrice *changement/ pas de changement* permettrait d'estimer l'omission (1-exactitude producteur) et la commission (1-exactitude utilisateur) des changements fournis par le SDC. La commission a été calculée grâce à l'étiquette que nous avons ajoutée *Pas de changement* et la classe *Changement* obtenue par agrégation des classes de référence *Résidentiel*, *Industriel*, *Infrastructure*, *Autre*, *Construction en cours* et *Destruction*. L'exactitude utilisateur obtenue est égale à 87%. Nous n'avons pas été en mesure d'estimer l'omission par manque de données de référence concernant les changements qui ont eu lieu sur le terrain. Cependant, à titre d'exemple, lors d'une *campagne flash* avec les experts de l'AuAT (Profil 5), 142 changements non détectés par le SDC ont été saisis par les experts en tant qu'alertes sous forme de polygones. Ceci montre la faiblesse du SDC à détecter la totalité des changements.

La Figure 19 montre un exemple d'alerte ayant la même classe d'usage selon le SDC et la référence calculée à partir de validations multiples (*i.e.* changement industriel). L'accord du contributeur est égal à 0,4.



Figure 19. Exemple montrant une alerte attribuée au type Industriel par le SDC et par la majorité des contributeurs. Extrait de l'outil Paysages avec deux références orthophotos : (a) 2016-Geoportail©IGN et (b) 2019-Geoportail©IGN ; issue de (Olteanu-Raimond et al., 2020)

La Figure 20 illustre l'impact de l'échelle dans la tâche de validation. Le SDC a détecté une construction résidentielle (polygone orange) tandis que la totalité des contributeurs ont attribué à

cette alerte l'étiquette *Pas de changement*. En analysant les deux orthophotos de 2016 et 2019 à une plus petite échelle que celle affichée dans l'outil de validation, nous pouvons remarquer une grande zone en construction, mais le changement détecté par le SDC est situé en dehors des constructions.



Figure 20. Exemple de changement type Résidentiel détecté par le SDC et considéré comme *Pas de changement* par la totalité des contributeurs. Extrait de l'outil Paysages avec deux références orthophotos : (a) 2016-Geoportail©IGN et (b) 2019-Geoportail©IGN ; issue de (Olteanu-Raimond et al., 2020)

Nous avons essayé de reproduire la validation et nous avons remarqué qu'à l'échelle choisie par défaut dans Paysages, et même en dé-zoomant, on ne remarque pas la construction dans la parcelle. A une échelle locale les contributeurs ont raison (le protocole est donc respecté) et le SDC a tort : le polygone détecté ne correspond pas à un changement. En revanche, à l'échelle de la parcelle il y a un changement, non détecté par le SDC ni remonté comme alerte par les contributeurs. Nous avons donc supposé que les contributeurs n'ont pas dé-zoomé pour voir toute la zone. Cela met en évidence la nécessité de déterminer l'échelle adaptée pour la tâche de validation.

Analyse des profils de contributeurs

L'implication de différents profils de contributeurs suscite de nombreuses questions de recherche telles que : quelles sont les connaissances nécessaires pour la validation de l'usage des nouvelles constructions ? Est-ce que la motivation des contributeurs a un impact sur la qualité de la validation ?

Pour mieux évaluer la capacité des différents profils à réaliser la tâche de validation, nous avons d'abord calculé la matrice de confusion entre les classes dérivées de chaque groupe de profils et la référence (*i.e.* le vote majoritaire). Au total, sept matrices de confusion ont été calculées. De la même manière que pour l'exactitude sémantique, seules les alertes ayant un accord des contributeurs supérieur à un seuil ont été sélectionnées pour l'analyse.

La Figure 21 illustre les exactitudes globales obtenues pour chaque profil. Par exemple, les exactitudes globales les plus importantes ont été obtenues pour le profil 4 (experts en recherche), le profil 5 (experts de l'AuAT) et le profil 3 (étudiants de l'ENSG volontaires), avec une exactitude respective de 93,2 %, 86,5 % et 85,2 %. L'exactitude la plus faible a été obtenue pour le profil 1, gr2 (étudiants de l'ENSG), le profil 2 (étudiants sans expertise en photo-interprétation) et le profil 6 (similaire à un profil de citoyen sans connaissance en géomatique). Nous avons été surpris par la différence entre les exactitudes obtenues pour les groupes 1 et 2 du profil 1, bien que ce soient des contributeurs ayant la même expertise. La différence pourrait être expliquée par un dysfonctionnement technique pendant la *campagne flash*. En effet, pendant de courtes périodes de temps, les flux WMS orthophoto 2019 n'étaient pas disponibles. Ainsi, dans certains cas, il a été difficile de valider le changement en utilisant uniquement Google Maps et les orthophotos de 2016, d'autant plus que nous avons remarqué que l'imagerie de Google Maps n'était pas à jour dans la

zone d'étude. Les exactitudes faibles pour le profil 2 et le profil 6 peuvent s'expliquer par le manque d'expertise en photo-interprétation et en classification d'usage. Enfin, nous pouvons observer une différence relativement importante entre le profil 1, gr1 (pour lequel la *campagne flash* faisait partie du programme d'études) et le profil 3 (pour lequel la *campagne flash* était une activité bénévole). Nous constatons que les étudiants de première année en début d'année, qui avaient peu d'expertise mais étaient bénévoles, ont mieux réalisé la tâche de validation que les étudiants de première année en milieu d'année, qui avaient suivi une formation.

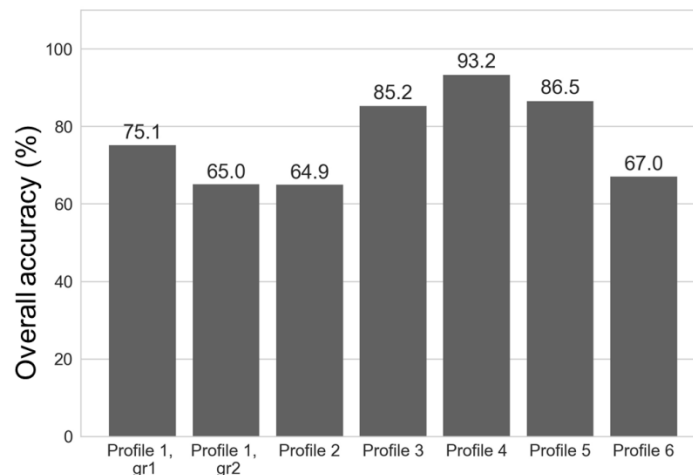


Figure 21. Exactitudes globales obtenues pour chaque profil en comparant leurs validations avec la référence.

Enfin, il nous a semblé important de savoir si l'expertise des contributeurs a un impact sur le choix des classes et si certaines classes peuvent produire plus de confusion que d'autres. Pour répondre à cette question, nous avons calculé des matrices de confusion en comparant les profils deux à deux. Même si des confusions de classes ont été remarquées pour la quasi-totalité des profils, le plus marquant a été la comparaison des profils 5 (experts AuAT) et 6 (qui est similaire à un profil de citoyen sans connaissance en géomatique). L'exactitude utilisateur pour les classes *Pas de changement* et *Résidentiel* est bonne (respectivement 90% et 86%). En revanche, de faibles exactitudes utilisateurs sont obtenues pour les classes *Construction en cours* et *Industriel*. Par exemple, 44 des alertes (36%) sont validées comme *Résidentiel* par le profil 5 et *Construction en cours* par le profil 6. Ceci montre que les experts ont analysé le contexte spatial même lorsque la construction n'était pas complètement terminée et ont estimé le type d'usage de la construction. De plus, nous avons appris que pendant la campagne impliquant le Profil 5, les contributeurs ont utilisé le flux d'OCS-GE disponible dans l'application Paysages pour interpréter le changement et qu'ils ont eu des échanges entre eux. Enfin, une différence importante concerne la classe *Inconnu* : 13% des alertes ont été validées comme *Inconnu* par le Profil 5 contre 0.02 % pour le Profil 6. Cela montre que les experts ont essayé de valider chaque alerte, même lorsqu'il était difficile de choisir un type précis.

Nous avons observé grâce à ce travail que les données d'occupation du sol sont très complexes et qu'il existe de nombreux problèmes qui restent à résoudre pour améliorer le processus de validation collaborative en vue de définir un jeu de données de référence et d'intégrer les alertes pour mettre à jour une base de données géographique d'autorité. Un premier verrou est de savoir quel est le format adapté pour une *campagne flash*. Par exemple, les experts LULC ont passé beaucoup plus de temps à interpréter une seule alerte ; en plus de leur expertise, ils ont utilisé les spécifications de l'OCS-GE et ont eu des discussions entre eux pour le choix final. Un comportement similaire, à l'exception de la consultation des spécifications, a été constaté pour la *campagne flash* pour laquelle un appel à une participation bénévole a été lancé aux étudiants. Un autre verrou concerne la conception des réponses. Une liste de huit classes a été jugée complexe par les citoyens et les étudiants sans expérience de l'ENSG, contrairement aux experts de l'AuAT qui ont estimé que le choix des classes n'était pas assez détaillé et qu'ils auraient souhaité davantage détailler la classe

Autre. Dans nos futurs travaux de recherche, nous aimerions tester la conception des réponses en fonction du profil des contributeurs (par exemple, une réponse simplifiée en changement / pas de changement pour les citoyens et les étudiants, et une réponse plus détaillée pour les experts en données de l'occupation du sol). Pour ces derniers, des discussions ont déjà été entamées afin de définir les classes d'occupation du sol nécessaires pour mieux répondre à leur besoins.

3.3 Proposition d'une méthodologie de qualification de l'information géographique volontaire issue de plateformes d'activités sportives

Avant-propos

Cette section présente une partie des travaux de la thèse de Stefan Ivanovic soutenue en 2018. La thèse, que j'ai encadrée, a été effectuée sous la direction de Sébastien Mustière et Thomas Devogele.

Cette section résume donc les travaux de la thèse qui portent sur l'évaluation et l'amélioration de la qualité des traces issues d'activités sportives :

Ivanovic, Stefan. 2018. 'Quality Based Approach for Updating Geographic Authoritative Datasets from Crowdsourced GPS Traces'. Thèse. Paris, France: Université Paris-Est.

Ivanovic, Stefan, Ana-Maria Olteanu-Raimond, Sébastien Mustière, and Thomas Devogele. 2019. 'A Filtering-Based Approach for Improving Crowdsourced GNSS Traces in a Data Update Context'. *ISPRS International Journal of Geo-Information* 8 (9): 380

La mise à jour des sentiers de montagne est une tâche très complexe et coûteuse en temps à cause de la nature très variable des sentiers, du changement saisonnier et de la difficulté d'utiliser les différentes sources d'images à cause du couvert végétal (*e.g.* en zone forestière les sentiers ne peuvent pas être vus sur les images). De plus, malgré l'importance des sentiers pour la réalisation des cartes de randonnée, la mise à jour des sentiers de la BDTOPPO® est moins prioritaire que le réseau routier principal. Dans ce contexte, la thèse de Stefan Ivanovic explore l'hypothèse selon laquelle l'IGV, malgré ses limites, est une source riche d'information qui peut être utile pour la mise à jour des sentiers du référentiel routier BDTOPPO® : mettre en évidence des objets géographiques manquants ou confirmer que certains objets géographiques existent toujours. Pour explorer cette piste, nous nous sommes focalisés sur l'usage de l'IGV représentant les traces GNSS issues d'activités sportives en zone de montagne.

Dans ce travail de recherche, une trace GNSS collectée à travers une activité physique (*e.g.* randonnée, course à pied, cyclisme) est représentée par un ensemble de points de coordonnées (lat, long), une estampille temporelle de type *timestamp* et une élévation. Dans la modélisation proposée par S. Ivanovic (voir la Figure 22), la trace représente la trajectoire d'un individu entre une origine (A) et une destination (B) et est caractérisée par deux types de comportement. Un comportement principal selon lequel l'individu se déplace en suivant un sentier et un comportement secondaire qui décrit le fait que l'individu effectue des activités en dehors du sentier (*e.g.* pause pique-nique, quitter le sentier pour voir un point de vue). Ce dernier type de comportement a été nommé comportement humain secondaire (CHS) et a été considéré dans notre contexte de recherche comme un bruit local de la trace, non pertinent pour reconstituer le tracé des sentiers.

D'autre part, comme évoqué en Section 3.1, de nombreux facteurs peuvent influencer la localisation d'un point de la trajectoire pendant le processus d'acquisition, générant ainsi un écart de position par rapport à la position réelle de l'individu. Ces erreurs ont été nommées des points de localisation aberrants (outliers).

Ainsi, un premier objectif a été de détecter et filtrer le comportement humain secondaire et les points aberrants afin d'améliorer la géométrie de la trace et de la rendre plus lisse et plus proche de la modélisation topographique d'un sentier. Le deuxième objectif a été de qualifier l'exactitude de la

position de chaque point composant la trace selon deux étiquettes : bonne exactitude et faible exactitude.

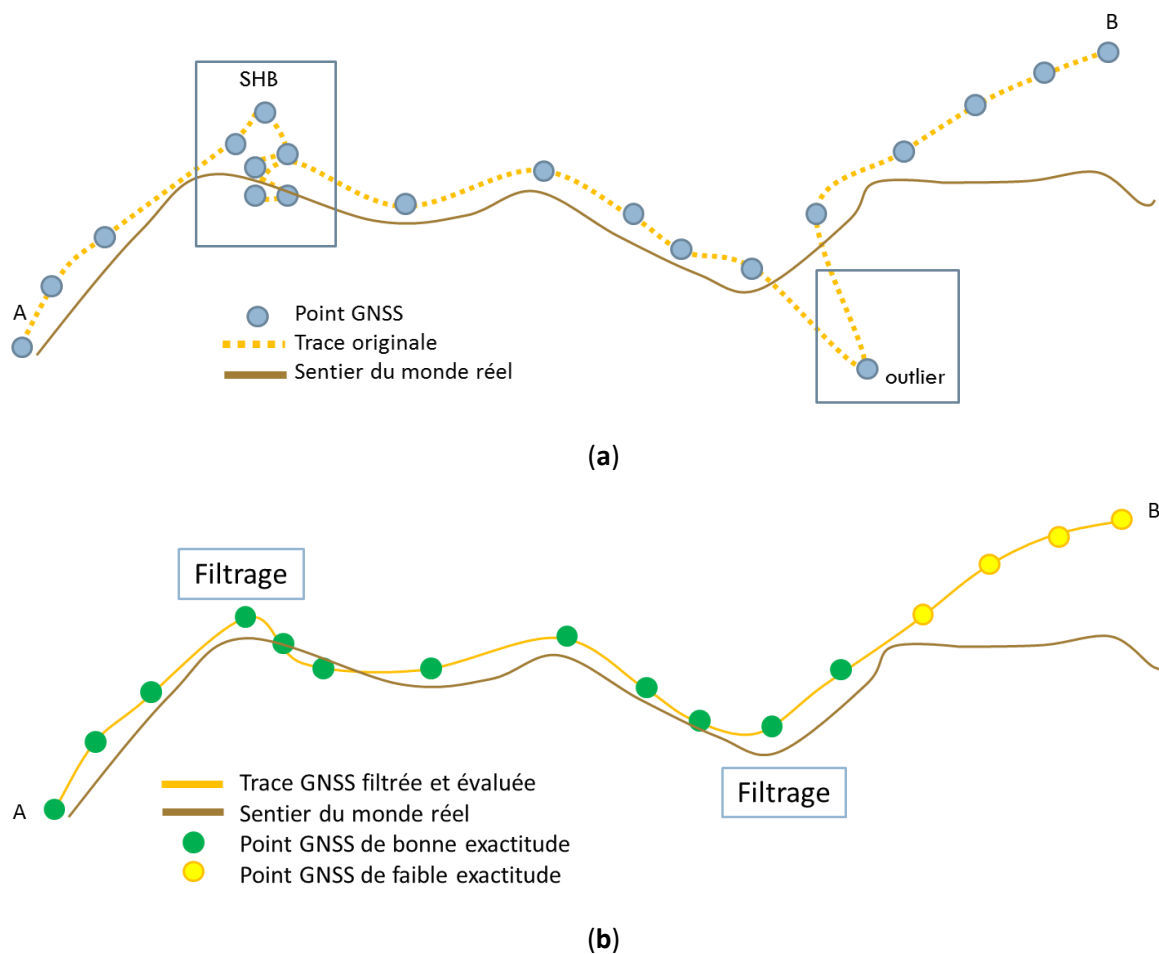


Figure 22. Exemple d'une trace d'un individu se déplaçant sur un sentier entre le point A et le point B et effectuant une activité de type pause pique-nique (SHB-comportement humain secondaire) : (a) trace originale; (b) résultat attendu dans notre approche. image adaptée de (S. Ivanovic et al., 2019)

3.3.1 Méthode pour la détection du comportement humain secondaire

La première étape de l'approche a été d'identifier le comportement humain secondaire pour chaque trace. Etant donné l'incomplétude thématique des traces collaboratives (*e.g.* absence d'estampille temporelle), les méthodes de détection de stations proposées dans la littérature ne sont pas adaptées (Alvares et al., 2007; Ester et al., 1996; Thierry et al., 2013). Par conséquent, S. Ivanovic a proposé une méthode de détection qui s'appuie uniquement sur les caractéristiques géométriques de la trace. Ainsi, le comportement humain secondaire est caractérisé par un changement de direction entre deux segments consécutifs et l'auto-intersection de la trace.

L'approche est définie de la manière suivante. Tout d'abord, les polygones déterminés par les parties de la trace qui s'auto-intersectent sont construits. Ensuite, pour chaque polygone, deux indicateurs sont calculés, à savoir l'aire et l'élongation. Enfin, les polygones dont l'aire et l'élongation sont inférieures à un seuil, choisi empiriquement, sont identifiés comme étant des polygones représentant les limites du comportement humain secondaire. Les points faisant partie de cette catégorie de polygones sont filtrés. Le processus se déroule de manière itérative jusqu'à l'élimination de toutes les auto-intersections. A la fin du processus, la trace est reconstituée en utilisant les points restants.

Précisons que l'élongation est définie comme le rapport entre l'aire du polygone et son plus petit polygone circulaire.

3.3.2 Méthode pour la détection des points GNSS aberrants

Un point aberrant a été défini comme un point dont les caractéristiques métriques et géométriques diffèrent sensiblement des caractéristiques des autres points composant la trace (Ivanovic, 2018). La méthode de détection des points aberrants est basée sur une méthode d'apprentissage automatique supervisée à base de règles. L'apprentissage supervisé à base de règles nous a semblé adapté dans notre cas car les règles apprises sont plus facilement interprétables et implémentables dans un langage de programmation et elles peuvent être modifiées ou améliorées, voir retirées si elles sont considérées comme incohérentes, ce qui n'est pas le cas des algorithmes de classification non symboliques.

La première étape consiste à définir un ensemble d'indicateurs intrinsèques (Tableau 3) et extrinsèques (Tableau 4). Au total quinze indicateurs ont été définis en prenant en compte les connaissances issues de la littérature concernant les facteurs qui influencent la position des points GNSS (Cain et al., 2005; N.J. DeCesare et al., 2005; D'Eon et Delparte, 2005).

Les indicateurs intrinsèques sont calculés localement à l'échelle de chaque point de la trace en utilisant des métriques provenant des traces elles-mêmes. Selon l'indicateur, deux ou trois points consécutifs ($i-1$, i , $i+1$) sont pris en compte. Dans le Tableau 3, D_{i-1} et D_i représentent respectivement la distance du point i au point précédent $i-1$, et la distance du point i au point suivant $i+1$. Idem pour les directions et les vitesses.

Les indicateurs extrinsèques, quant à eux, sont dérivés à partir du contexte spatial des points composant la trace tel que le type et la densité du couvert végétal, la pente ou la proximité d'obstacles (*e.g.* falaises et bâtiments, lacs).

Tableau 3. Définition des indicateurs intrinsèques ; issu de (Ivanovic et al., 2019)

Indicateur	Description	Formule
AngleMean	Valeur moyenne des 3 directions consécutives	$(\alpha_{i-1} + \alpha_i + \alpha_{i+1})/3$
DistDiffN	Distance normalisée entre 3 points consécutifs	$(D_{i-1} - D_i)/(D_{i-1} + D_i)$
DistDiffMed	Distance entre trois points consécutifs normalisée par la distance médiane de la trace	$\frac{D_{i-1} + D_i}{2\text{Median}(\text{Trace})}$
DistMean	Distance moyenne entre trois points consécutifs	$(D_{i-1} + D_i)/2$
SpeedDiffN	Vitesse normalisée	$(V_{i-1} - V_i)/(V_{i-1} + V_i)$
SpeedMean	Vitesse moyenne entre trois points consécutifs	$(V_{i-1} + V_i)/2$
SpeedRate	Taux de vitesse	$(V_{i-1} + V_i)/V_i$
DiffElev	Max de différence d'altitude	$\max Z_{i+1} - Z_i, Z_i - Z_{i-1} $

Les indicateurs extrinsèques (Tableau 4) prennent en compte les facteurs influençant l'acquisition des points GNSS comme évoqués en Section 3.1. Trois types de données géographiques d'autorité ont été utilisés, à savoir la couverture des sols, les MNT et les données topographiques de la BDTOPO® de l'IGN.

La deuxième étape consiste à générer des exemples d'apprentissage sur une zone représentative de notre zone d'étude ; chaque point a été classifié manuellement selon deux étiquettes : aberrant et

non aberrant. La troisième étape consiste à appliquer l'algorithme RIPPER¹⁵ (Cohen, 1995) sur la partie restante de la zone d'étude. Enfin, les règles apprises ont été implémentées et appliquées sur les traces de notre zone d'étude. Les points détectés comme aberrants ont été filtrés, puis les traces reconstruites.

Tableau 4. Définition des indicateurs extrinsèques; issue de (S. Ivanovic et al., 2019)

Indicateurs	Description	Formule
DiffElevDTM	Différence entre l'élévation (GNSS et MNT)	$ Z_{MNT} - Z_{GNSS} $
Slope	Gradient de pente	$\tan(\theta)$, $-90^\circ < \theta < 90^\circ$
Obstacles	Proximité des obstacles	Vrai si proche d'un obstacle, sinon faux
Curvature	Convexité de la pente	$1/R$
Vegetation	Type de forêt	f (couverture des sols)
CanopyCover	A l'intérieur de la forêt	f (couverture des sols), boolean
InBuildingWater	A l'intérieur d'un bâtiment ou d'un point d'eau	f (données topographiques), boolean

3.3.3 Méthode pour l'évaluation de l'exactitude des points GNSS de la trace

Cette section décrit une approche basée sur l'apprentissage supervisé à base de règles permettant d'évaluer l'exactitude des points GNSS de la trace selon deux étiquettes : bonne et faible (voir la Figure 22). L'approche est similaire à celle proposée pour détecter les points aberrants et les mêmes indicateurs intrinsèques et extrinsèques sont utilisés pour cette classification.

Pour évaluer l'exactitude des points GNSS, de nouveaux exemples d'apprentissage ont été générés. La définition des exemples étant une tâche coûteuse en temps, S. Ivanovic a proposé trois approches : manuelle, automatique et semi-automatique. Dans l'approche manuelle, l'exactitude des points est évaluée manuellement pour minimiser les faux positifs ou négatifs. Elle est guidée par des cartes Scan 25 et des orthophotos. Dans l'approche automatique, les points sont automatiquement classés en bonne ou faible exactitude en fonction de leur distance par rapport à la route la plus proche du réseau routier de la BDTOPO®. Notons que seuls les types de route qui peuvent être pratiqués à pied ou à vélo ont été sélectionnés. Cette approche permet d'acquérir de nombreux exemples d'apprentissage ; toutefois, il convient de noter sa limite dans le cas où un point GNSS de bonne exactitude est classé en faible exactitude car le sentier n'est pas représenté dans la BDTOPO®. Enfin, dans l'approche semi-automatique, les points sont automatiquement classés en fonction de leur distance par rapport à la route la plus proche, mais les exemples sont ensuite filtrés manuellement en prenant en compte l'hypothèse selon laquelle la route n'est pas représentée dans la BDTOPO®. Cette dernière approche est un compromis entre la qualité des exemples générés et le temps alloué à cette tâche.

Il convient de noter que cette approche a été appliquée aux traces filtrées grâce aux méthodes décrites ci-dessus, à savoir la détection des comportements humains secondaires et des points aberrants.

3.3.4 Application et résultats

La méthodologie de qualification des traces issues de la foule a été appliquée sur 437 traces (9773 km) issues d'activités sportives dans les Vosges venant de différents sites web de partage tel que randoGPS, tracesGPS, visuGPS, et VTTour (voir la Figure 23).

¹⁵ Repeated Incremental Pruning to Produce Error Reduction

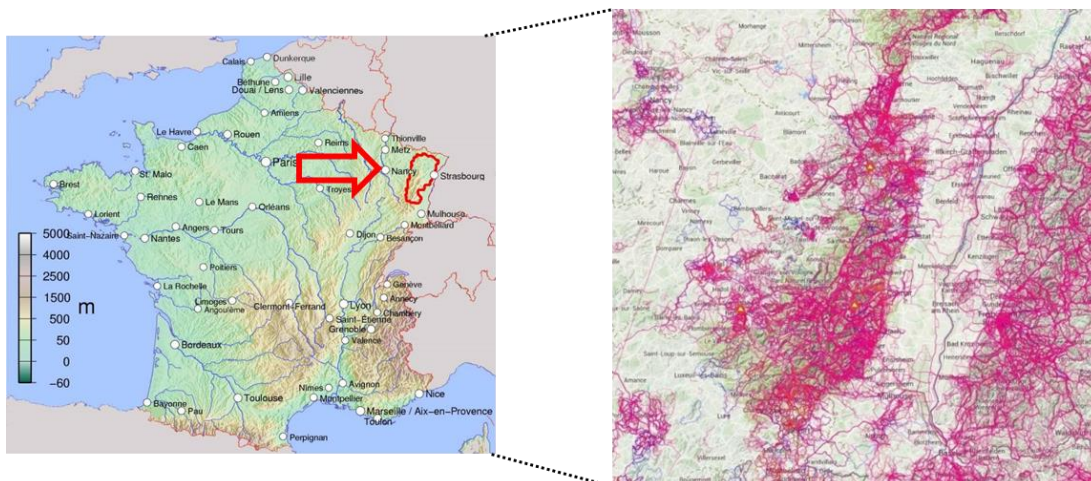


Figure 23. Zone d'étude pour la qualification des traces GNSS; issue de (Ivanovic 2018)

Etant donné le manque de protocoles de collecte, aucune métadonnée (*e.g.* type de capteur, constellation, type d'activité, cadence temporelle d'acquisition) n'est disponible. Toutefois, pour certaines traces la date d'acquisition de la trace est présente. L'incomplétude thématique est caractérisée par l'absence de valeur pour l'estampille temporelle (*e.g.* 36,3 % n'ont pas de valeur remplie) ou pour l'élévation (2,2 %). A l'échelle de la trace, 157 traces (35,9%) sont dépourvues de la totalité des estampilles temporelles, tandis que 287 d'entre elles (65,6%) ont au moins une estampille temporelle manquante. L'absence des estampilles temporelles est une vraie limite des traces, cette information étant nécessaire aux calculs de vitesse ou d'accélération. Des hétérogénéités spatio-temporelles très fortes ont été remarquées à la fois entre les traces et à l'intérieur d'une même trace. Par exemple, moins de 7% des traces ont une distribution normale pour les indicateurs de vitesse et aucune trace n'a de distribution normale pour les indicateurs de distance. De plus, différentes distributions de valeurs de vitesse entre les traces ont été remarquées (*i.e.* asymétriques à droite ou à gauche, fort ou faible aplatissement).

Détection du comportement humain secondaire (CHS)

La méthode de détection des points représentant le comportement humain secondaire a permis de détecter 11 746 points (4% du total) comme faisant partie de la catégorie CHS. Une validation détaillée et quantitative des résultats a été effectuée sur 14 traces sélectionnées au hasard représentant 265 km, et une précision de 98% et un rappel de 93% ont été obtenus. La Figure 24 illustre un exemple de trace avant et après le filtrage du CHS.

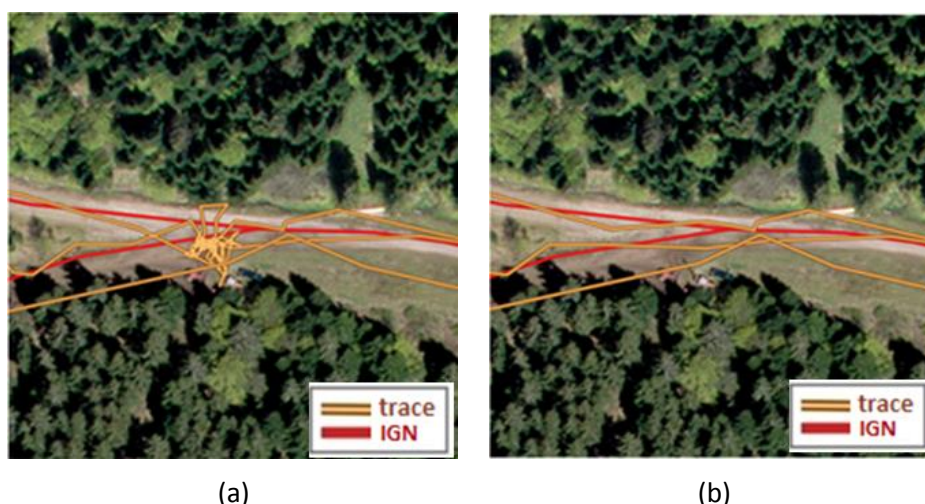


Figure 24. Exemple de détection et filtrage du comportement humain secondaire ; (a) trace avant filtrage du CHS et (b) après le filtrage ; issue de (Ivanovic et al. 2019)

Détection des points aberrants

Dans la zone d'échantillonnage sélectionnée, 2342 points ont été sélectionnés et étiquetés manuellement comme points aberrants ou non-aberrants. Parmi ces points, 35 (3 %) ont été étiquetés comme aberrants et 2 265 (97 %) comme non-aberrants. Il est évident que le jeu d'apprentissage est déséquilibré (un point aberrant pour 66 points non-aberrants). Pour réduire ce biais, S. Ivanovic a appliqué une des approches proposées dans la littérature pour faire face aux jeux de données déséquilibrés (Japkowicz, 2000). La méthode consiste à ajouter des points aberrants en dehors de la zone d'apprentissage. Au total, 77 nouveaux exemples ont été ajoutés. La représentativité statistique des exemples a été perdue, mais d'autres exemples de points aberrants ont été obtenus avec un effort limité.

Une fois le jeu de données d'apprentissage généré, l'algorithme RIPPER a été appliqué et une validation croisée a permis de mesurer la qualité des résultats. Ainsi, une précision et un rappel de 0.78 ont été obtenus. Les résultats ont confirmé la performance de l'approche et des règles apprises, malgré le manque de métadonnées et l'incomplétude de l'information thématique. Cinq règles ont été apprises pour détecter les points aberrants (voir Figure 24). Force est de constater que seuls les indicateurs intrinsèques sont présents dans les règles apprises, ce qui permet de conclure que seules les données elles-mêmes permettent de détecter les points aberrants. Ceci est un résultat très intéressant, car cela veut dire que l'exactitude de la trace peut être améliorée sans utiliser d'information géographique de référence.

Tableau 5. Règles apprises pour la détections des points aberrants ; issu de (Ivanovic et al., 2019)

Règle apprise	Description
Règle 1	IF DistDiffMed ≥ 1.05 AND AngleMean $\geq 87.54 \rightarrow$ outlier
Règle 2	IF AngleMean ≥ 71.25 AND SpeedRate $\geq 1.50 \rightarrow$ outlier
Règle 3	IF AngleMean ≥ 74.80 AND DistDiffN $\leq 0.21 \rightarrow$ outlier
Règle 4	IF AngleMean ≥ 83.15 AND SpeedRate $\leq 0.85 \rightarrow$ outlier
Règle 5	IF AngleMean ≥ 56.43 AND DistMean $\geq 8847.31 \rightarrow$ outlier

La Figure 25 (a) illustre un exemple de points GPS aberrants (en jaune) détectés à raison par la règle 1 qui permet de détecter des changements de direction brusque (90°) et par des distances entre deux points consécutifs plus grandes que la valeur des distances médianes de la trace, qui est un indicateur de la résolution géométrique de la trace. L'exemple de la Figure 25 (b) montre un faux positif. Les points aberrants en jaune ont été détectés à tort comme aberrants. L'analyse des résultats a permis de voir que les traces à faible résolution spatiale et à forte sinuosité contenaient la plupart des faux positifs. Il s'agit principalement d'un effet secondaire causé par l'indicateur AngleMean.

Cependant, si les traces à forte sinuosité ont une bonne résolution géométrique des traces (voir Figure 25 (b)), l'effet secondaire de l'indicateur d'angle n'existe plus, étant minimisé par le fait qu'un ensemble de points consécutifs ont les mêmes caractéristiques.

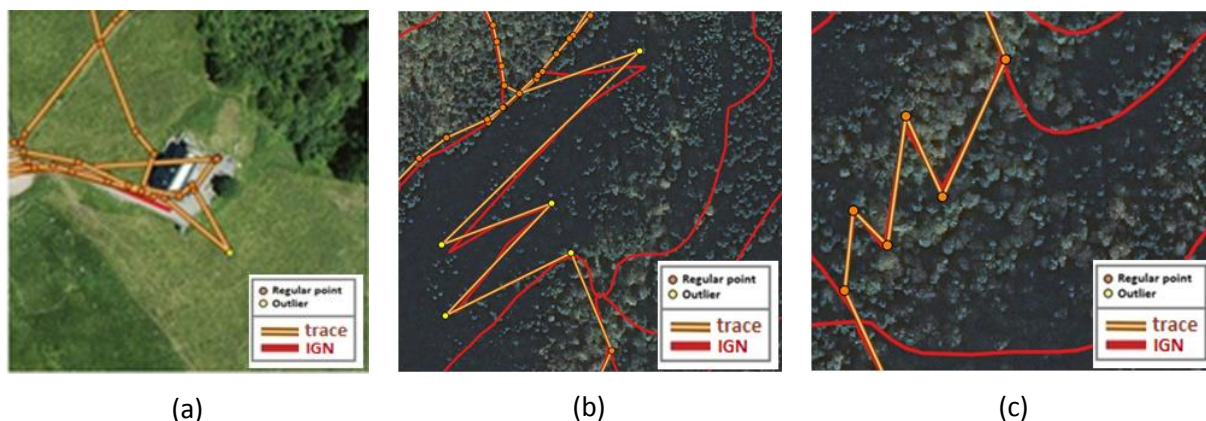


Figure 25. Exemples de détection des points aberrants : (a) bonne détection, (b) faux positif et (c) bonne détection; issues de (Ivanovic et al. 2019)

L'analyse des faux négatifs a montré que ces derniers sont en général dus à la sensibilité des seuils appris. Comme le montre le Tableau 5, les seuils appris sont très précis, ce qui génère bien évidemment une sensibilité mais surtout, ils montrent la difficulté à définir des seuils pour la détection des points aberrants dans les traces provenant de la foule.

Les traces ont été reconstruites après le filtrage des points aberrants. Ainsi, à la fin de cette étape, les traces GNSS issues des activités sportives ont un tracé géométrique plus lisse, les micro-déplacements liés à l'activité sportive et les erreurs d'acquisition étant supprimées.

Qualification de l'exactitude des points GNSS de la trace

La méthode de qualification de l'exactitude a été appliquée sur les traces filtrées. Comme évoqué dans la Section 3.3.3, les trois approches ont été appliquées pour générer les exemples d'apprentissage, ensuite le même algorithme RIPPER et les mêmes indicateurs intrinsèques et extrinsèques sont utilisés pour générer des règles de classification. Les résultats obtenus sont synthétisés dans le Tableau 6. La précision et le rappel obtenus par validation croisée montrent que les résultats sont moins satisfaisants que ceux obtenus pour la détection des points aberrants. Cela montre à quel point la tâche est difficile. Une explication est certainement que les caractéristiques des points aberrants sont clairement différentes de celles des points non-aberrants, alors que les caractéristiques des points de faible exactitude ne sont pas significativement différentes de celles des points de bonne exactitude. Cependant, l'approche basée sur l'apprentissage supervisé permet de détecter environ la moitié des points de faible exactitude.

Tableau 6. Résultats de la qualification de l'exactitude des points GNSS

Approche	% de points par catégorie	Nombre règles apprises	Précision	Rappel
Manuelle (2008 points)	76% bonne exactitude	7	0.48	0.62
	24% faible exactitude			
Automatique (14721 points)	73% bonne exactitude	23	0.4	0.7
	27% faible exactitude			
Semi-automatique (7745 points)	70% bonne exactitude ;	27	0.79	0.62
	30% faible exactitude			

L'analyse approfondie des règles apprises, ainsi que la vérification visuelle, mettent en évidence les principaux avantages et inconvénients des trois approches pour la génération des exemples d'apprentissage. L'approche manuelle donne des règles simples et génériques, ce qui est important. C'est la méthode qui prend le plus de temps, mais les erreurs d'échantillonnage sont minimisées. Même si la précision et le rappel ne sont pas élevés, ils sont satisfaisants, compte tenu de l'hétérogénéité des données utilisées.

L'approche semi-automatique permet un compromis entre le gain de temps et l'obtention d'exemples de qualité : elle donne les meilleurs scores. Néanmoins, la vérification visuelle montre que les règles de détection des points de faible exactitude dépendent des exemples choisis. Il se peut que les traces utilisées dans l'échantillonnage ne soient pas si hétérogènes (par exemple, enregistrées avec des appareils similaires, dans des conditions similaires). Par conséquent, les règles ont des caractéristiques locales et ne fonctionnent pas bien en dehors de la zone d'échantillonnage.

Enfin, l'approche automatique fournit un grand nombre d'exemples et permet de gagner du temps. Toutefois, elle inclut des points correspondant à des routes non représentées dans l'ensemble des données faisant autorité, qui peuvent être classés à tort comme des points de faible exactitude. En effet, la précision de cette approche est la plus faible. C'est pourquoi l'approche manuelle a été préférée dans ce travail, et seules les règles apprises à travers cette approche seront appliquées (voir Tableau 7).

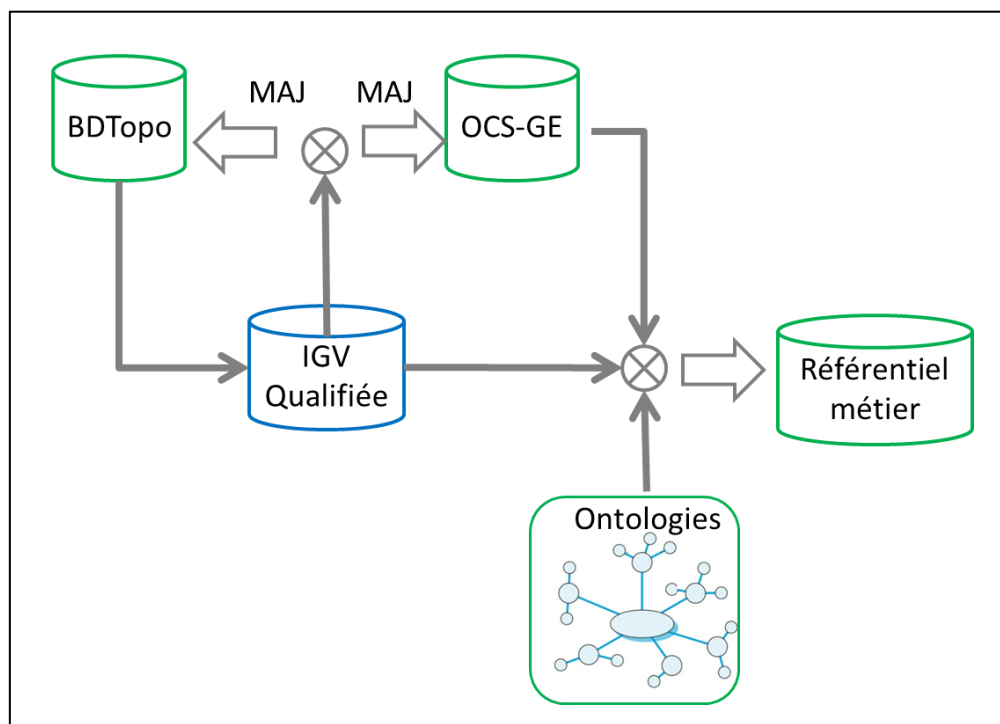
Un résultat intéressant qui mérite d'être souligné est la présence des indicateurs extrinsèques dans les règles apprises. Il s'agit des indicateurs se basant sur la couverture de canopée, la pente et la différence d'élévation par rapport au MNT. Ce résultat conforte les connaissances issues de la littérature, à savoir que la couverture végétale et le relief sont les facteurs qui influencent le plus l'exactitude des localisations lors de l'acquisition.

Même si les résultats ne sont pas parfaits, notre hypothèse est que la détection des points de faible exactitude est une source qui pourrait être exploitée pour la mise à jour des données d'autorité, à condition que cette imprécision soit prise en compte dans le processus de décision.

Tableau 7. Règles apprises pour la qualification de l'exactitude de points GNSS ; issu de (Ivanovic, 2018)

Règle	Description
Règle 1	IF AngleMean>45.05° and CanopyCover= YES and Slope>19.2 → faible exactitude
Règle 2	IF AngleMean ≥ 23.45° and SpeedMean ≤ 1.21m/s and DiffElevDTM ≥ 14m → faible exactitude
Règle 3	IF AngleMean >= 42.63° and DistDiffMed <= 0.62 m → faible exactitude
Règle 4	DistMean ≥ 45.37m and DiffElevDTM ≤ 4.83m and CanopyCover = YES and Slope ≤ 17.55 and DistDiffMed ≤ 1.81m → faible exactitude
Règle 5	IF AngleMean ≥ 21.14° and DistMean ≥ 116.34m → faible exactitude
Règle 6	IF Slope ≥ 25.18 and DistMean ≤ 11.16m → faible exactitude
Règle 7	IF CanopyCover= YES and Slope>31.2 and AngleMean>21 → faible exactitude

Chapitre 4. Intégration de l'information géographique hétérogène



Partie de la démarche globale (Figure 1) décrite dans ce chapitre

Ce chapitre décrit les travaux de recherche qui permettent de répondre au quatrième objectif scientifique énoncé dans le Chapitre 1, à savoir **(3)** Intégrer différentes sources d'information géographique volontaire et institutionnelle en prenant en compte leur hétérogénéité. L'hypothèse de recherche explorée dans ce chapitre est que, malgré les limites et l'hétérogénéité de l'IGV, elle est une source riche pour la mise à jour de référentiels d'autorité et pour la constitution d'un référentiel métier.

4.1 Problématiques et positionnement

Le terme d'intégration est utilisé dans ce chapitre dans un sens large. Il englobe l'analyse de différentes sources d'information géographique afin de déterminer leurs similitudes, leurs différences, et leurs complémentarités, d'une part, et leurs combinaisons afin de répondre à un besoin donné, d'autre part. Le positionnement de mes recherches pour l'intégration de l'IGV est guidé par les objectifs opérationnels évoqués en Introduction.

Une des questions principales que l'on se pose est : Quel est le potentiel de l'IGV pour la mise à jour de référentiels d'autorité ? Cette question implique deux questions sous-jacentes : (1) L'IGV permettra-t-elle d'améliorer le processus de suivi des changements ? (2) Ou au contraire, pourrions-nous aller jusqu'à la mise à jour des référentiels ? La première question implique que l'IGV soit utilisée pour lever des alertes de mise à jour et que la mise à jour du référentiel soit réalisée en suivant le processus de mise à jour traditionnel. La deuxième question implique la proposition d'un processus de mise à jour intégrant l'IGV. Dans ce cas, un verrou à lever est la qualification des résultats et le rôle de l'opérateur dans le processus.

La constitution d'un référentiel métier à partir de plusieurs sources d'IGV soulève plusieurs questions concernant la gestion des formats et des sémantiques hétérogènes. Dans notre contexte applicatif (*i.e.* l'aide à la localisation des victimes en montagne), gérer les redondances s'avère une tâche primordiale pour la détermination d'une zone de localisation pour la victime, étant donné les indices fournis par la victime ou le requérant.

Besoin de mise à jour du référentiel routier de la BDTOPO®

Les moyens traditionnels de mise à jour des référentiels topographiques faisant autorité reposent globalement sur des enquêtes de terrain et des données de télédétection. Dans le cas des enquêtes de terrain, elles sont généralement effectuées sur la base de nouvelles informations provenant des autorités locales (*e.g.* la publication officielle de la construction d'un nouveau lotissement) de changements détectés à distance (*e.g.* à partir d'images). Une fois les alertes identifiées, les opérateurs chargés de la mise à jour effectuent les changements soit en allant sur le terrain, soit par stéréo-restitution. La mise à jour est une tâche très coûteuse. Le réseau routier est le thème ayant le plus petit écart de temps entre deux publications. L'actualité de certains type de routes tels que autoroute, départementale, nationale peut être de 6 mois, tandis que pour d'autres types de routes tels que les sentiers, elle peut aller jusqu'à 3-4 ans (Spécifications BDTOPO, 2019). Le type *sentier*¹⁶ est un type particulièrement difficile à mettre à jour en raison de la nature intermittente des sentiers (*e.g.* le passage fréquent dans un espace ouvert peut créer un sentier et au contraire un sentier peut disparaître s'il n'y a plus de passage) et de leur localisation dans des paysages parfois difficiles pour un suivi régulier (*e.g.* la forêt, la haute montagne, le bord de mer). L'utilisation de techniques de stéréo-restitution est difficile dans les zones montagneuses où la végétation dense peut limiter la visibilité depuis le ciel. Les enquêtes sur le terrain sont très coûteuses en raison de la vaste zone ouverte qui doit être couverte. Bien que les sentiers soient considérés comme moins prioritaires pour la mise à jour, ils sont nécessaires pour répondre à des besoins cartographiques et touristiques. Dans ce contexte, nous souhaitons améliorer l'actualité des sentiers en utilisant l'IGV.

L'utilisation de l'IGV à des fins de mise à jour ou d'enrichissement a été envisagée dans de nombreux travaux de recherche. Al-Bakri et Fairbairn (2012) concluent que malgré la richesse de la base de données OSM, il existe d'importantes incompatibilités, en particulier au niveau de l'exactitude, qui rendent difficile la mise à jour ou l'enrichissement du référentiel de l'Ordnance Survey¹⁷. D'autre part, Zielstra et Hochmair (2011) ont trouvé que le résultat d'intégration de données OSM avec les données de TeleAtlas et NAVTEQ est tout-à-fait satisfaisant. Cependant, cela fait référence à un

¹⁶ Les sentiers sont des « chemins étroits ne permettant pas le passage de véhicules », (Spécifications BDTOPO, 2019)

¹⁷ L'Agence Nationale de Cartographie de Grande Bretagne : <https://www.ordnancesurvey.co.uk/>

contexte spécifique où les sentiers pédestres OSM sont considérés comme utiles pour améliorer l'accessibilité aux stations de bus et de métro dans les villes américaines et allemandes. Van Winden et al. (2016) ont utilisé des traces issues de la foule pour mettre à jour l'information thématique du réseau routier d'autorité néerlandaise (e.g. les routes à sens unique ou à double sens). Liu et al. (2015) ont proposé une approche pour détecter les nouvelles routes à partir de la base OSM pour mettre à jour un réseau routier d'autorité, ce qui nécessite de lever les incompatibilités entre les données.

Comme dans les travaux ci-dessus, notre objectif est d'utiliser l'IGV et plus précisément les traces GNSS issues des activités sportives pour identifier des mises à jour du type *sentier*. Notre piste de recherche a été d'abord d'améliorer la qualité du tracé géométrique de la trace GNSS et d'estimer l'exactitude des points GNSS (voir la Section 3.3), puis de lever des alertes en identifiant des sentiers manquants en s'appuyant sur des outils d'appariement de données. Enfin, les alertes sont qualifiées en combinant différents critères, dont l'estimation de l'exactitude des points GNSS.

Besoin de mise à jour du référentiel d'occupation du sol de l'OCS-GE®

La mise à jour de l'occupation du sol s'appuie traditionnellement sur des informations multi-temporelles. Nous distinguons deux types d'approche : (1) par comparaison post-classification et (2) par détection des changements appliqués aux images. La comparaison post-classification est une approche très populaire qui consiste à classer les images à deux temporalités différentes avec les mêmes algorithmes, puis à comparer les deux classifications obtenues (Domenech et Mallet, 2014; Gressin et al., 2014). Les approches basées sur la détection du changement reposent pour la plupart sur l'analyse du vecteur de changement qui peut être appliqué aux séries chronologiques d'images contenant plus de deux temporalités (Bovolo et Bruzzone, 2007; Liu et al., 2015). Cette approche consiste à calculer les différences entre les pixels d'images à différentes époques afin d'obtenir les vecteurs de changement entre eux.

Ces dernières années, l'IGV est apparue également comme une alternative pour produire ou mettre à jour l'occupation du sol (Estima et Painho, 2013; Jokar Arsanjani et al., 2013; Fonte et al., 2017b) ou enrichir une occupation du sol existante (Fonte et al., 2017a; Schultz et al., 2017). Un travail pionnier dans la dérivation de l'occupation du sol à partir de l'IGV a été entrepris par (Estima et Painho, 2013) qui ont proposé une méthode pour dériver les usages du sol à partir de données OSM en utilisant une approche basée sur un arbre de décision hiérarchique. Des recherches similaires ont été proposées par Jokar Arsanjani et al., (2013) et Fonte et al., (2017b) pour dériver des données vectorielles représentant la couverture ou l'usage du sol. Bien que de nombreux progrès aient été réalisés, la classification d'usage du sol utilisée est grossière. Par exemple, les classes d'usage de type industriel, commercial, public, militaire, résidentiel et de transport ont été regroupées en une seule classe d'usage (correspondant au niveau 2 de la nomenclature d'Urban Atlas). De plus, l'utilisation de l'IGV provenant d'OSM pour produire une occupation du sol millésimée pose problème. En effet, mesurer l'évolution entre deux millésimes est un défi car tout changement de géométrie ou d'attribut d'un objet géographique ne représente pas forcément un changement sur le terrain mais peut être dû à une correction d'erreur. Dans OSM, il n'est pas possible de distinguer une mise à niveau (correction d'erreur) d'une mise à jour (changement sur le terrain).

Concernant la mise à jour de la couverture du sol, Fonte et al., (2017a) ont étudié le potentiel des données OSM pour mettre à jour et améliorer l'occupation du sol GlobeLand30, qui fait autorité. En se basant sur les approches décrites ci-dessus, ils ont d'abord rastérisé les données d'occupation du sol dérivées d'OSM en utilisant la même résolution que le produit GlobeLand30. Ensuite, une fois les pixels homologues de GlobeLand30 et d'OSM identifiés, ils ont effectué la mise à jour du pixel GlobeLand30 en transférant la classe de couverture du raster dérivé d'OSM, si cette dernière est disponible. La nomenclature de GlobeLand30 a été améliorée grâce à la nomenclature plus détaillée de l'occupation du sol dérivée d'OSM, en particulier dans les zones urbaines.

Comme pour les travaux proposés par Fonte et al., (2017a), notre objectif est d'enrichir et mettre à jour l'occupation du sol faisant autorité. Concernant l'enrichissement de l'usage, notre travail se concentre sur les zones bâties afin de distinguer les zones résidentielles, commerciales et industrielles. Notre **hypothèse est que les alertes de changements provenant de différentes sources d'IGV peuvent être utilisées pour mettre à jour l'OCS-GE**. L'usage des alertes pose plusieurs problèmes. Premièrement, il est nécessaire de distinguer les alertes significatives et non significatives pour l'occupation du sol. Par exemple, une nouvelle construction dans une zone agricole peut changer la classe d'usage ou de couverture de l'objet géographique, alors qu'une nouvelle construction n'a pas d'effet sur la couverture de l'objet géographique car elle était déjà de type bâti. Deuxièmement, lorsque différentes sources d'IGV sont utilisées, les mêmes objets géographiques de l'OCS-GE peuvent faire l'objet de plusieurs alertes de mises à jour ce qui permet d'obtenir des informations plus fiables ou complémentaires, mais aussi des informations potentiellement contradictoires. Notre piste de recherche est de prendre en compte le millésime actuel de l'OCS-GE et de combiner l'IGV disponible pour tirer parti de ces informations complémentaires en utilisant une approche de fusion d'informations basée sur la théorie de Dempster-Shafer. Cette dernière a été déjà utilisée pour des problématiques d'occupation du sol (Comber et al., 2004 a; Comber et al., 2004 b) et s'inscrit dans la continuité de mes travaux de thèse (Olteanu, 2008). Cette hypothèse a été explorée dans le cadre du post-doctorat de Lanfa Liu et décrite dans la Section 4.3.

Besoin de constituer un référentiel métier pour la localisation des victimes en montagne

Ce besoin s'inscrit dans le cadre du projet ANR Choucas qui vise à améliorer la phase de localisation des victimes en montagne menée par les secouristes (voir l'Annexe 3). Ne connaissant pas sa position exacte, la victime décrit son environnement spatial à travers des indices tels que : «Je vois le sommet du Mont Blanc», «Je suis près d'un refuge», ou en décrivant son itinéraire : «Je suis sur le GR54», «Je marche sur un chemin qui longe un lac». L'usage des objets de repère (sommet du Mont blanc, refuge) ou d'itinéraire (GR54) permettant à un humain de mieux s'orienter dans un espace qui ne lui est pas familier est bien connu dans la littérature (Lynch, 1960; Egorova et al., 2015; Zhou et al., 2017). Ce type de description qualitative qui consiste à localiser un objet ou une personne en s'appuyant sur des relations de localisation (voir, près de, sur, longer) et des objets de référence (e.g. sommet, refuge, GR, lac) est connu dans le domaine de la géomatique sous le nom de géoréférencement indirect (Hill et Zheng, 1999). Ceci est cohérent avec le concept de géographie naïve proposé par Egenhofer et Mark, (1995) dont l'objectif est de modéliser comment les individus perçoivent l'espace.

Pour passer d'une localisation relative à une localisation absolue définie par des coordonnées géographiques, une première nécessité est d'avoir accès à l'information géographique représentant des objets de repère et des itinéraires. Traditionnellement, les objets de repère sont représentés dans des référentiels issus des institutions telles que les ANC ou dans les gazetiers (e.g. Wikipedia, GeoNames). Ces derniers sont pour la plupart produits en mode collaboratif ou proviennent de sources d'IG externes, y compris celles faisant autorité (e.g. des organismes public locaux). Nous avons vu en Section 3.3 que de plus en plus de sites web collaboratifs offrent la possibilité de partager et de télécharger des traces GNSS. Certaines sont modélisées sous forme d'itinéraires, c'est-à-dire des sentiers ou des parcours à suivre et sont portés par une infrastructure de réseau physique composée de routes et de sentiers piétonniers (Teulade-Denantes et al., 2015). Il existe également des sites web qui fournissent des objets de repères (e.g. refuges.info) ; certains sites partagent à la fois des objets de repère et des itinéraires (e.g. campToCamp). En plus de l'hétérogénéité de l'IGV collectée et partagée par les contributeurs dont nous avons discuté dans le Chapitre 3, nous sommes confrontés aussi à l'hétérogénéité issues des modèles de données très différents d'un site web à un

autre (e.g. types d'objets de repère différents, cotations de difficulté différentes, plusieurs noms pour le même itinéraire, modèle trace versus itinéraire). Un premier besoin sous-jacent a été **d'évaluer les différentes sources en fonction de leur modèle et de leur complémentarité**.

De plus, connaître les objets de repère à proximité d'un itinéraire ou les itinéraires passant à proximité de certains objets de repère est une connaissance qui s'avère très utile pour les secours en montagne pour valider ou invalider une hypothèse de localisation de la victime ou bien pour poser une nouvelle question. Or, dans ce cas aussi, il y a une hétérogénéité entre les différentes sources d'IGV, certaines ne contenant que les objets de repère, d'autres que les itinéraires, d'autres encore contenant les deux à la fois. Pour cette dernière catégorie, dans certaines sources il y a déjà une relation entre les objets de repère et les itinéraires tandis que dans d'autres, la relation n'est pas connue. Lorsqu'il y a une relation, elle n'est pas nommée. Dans ce contexte, un deuxième besoin sous-jacent est d'identifier **les relations qui existent entre les objets de repère et les itinéraires d'une part, et de catégoriser ces relations (e.g. visibilité, proximité, contourner), d'autre part**. Cette modélisation est différente de celles proposées dans un contexte de navigation (Denis, 1997) ou de géocodage (Moncla et al., 2014), où les objets de repère font partie de l'itinéraire et sont considérés comme des points de passage. Enfin, pour faciliter l'interrogation des objets de repère et des itinéraires ainsi que des relations entre eux, un troisième besoin sous-jacent porte sur **l'intégration multi-sources et multi-objets**.

Pour répondre à ce besoin de constitution de référentiels métiers et aux besoins sous-jacents, nous suivons trois pistes de recherche : Une **première piste de recherche consiste à se baser sur la modélisation des connaissances métier** relatives aux types objets de repère et itinéraires et aux relations de localisation sous forme d'ontologies pour faciliter l'intégration, l'interrogation et la détermination de la zone de localisation de la victime selon un indice donné. Nous nous appuyons sur les travaux de recherche de (Bateman et al., 2010; Mustière et al., 2011b) pour les ontologies et sur les outils d'appariement de données géographiques afin de générer de nouvelles connaissances autour de différentes sources d'IG (Olteanu-Raimond et al., 2015) et d'estimer leur complémentarité et leur ressemblance (Derungs et S. Purves, 2013; Acheson et al., 2017).

Cette piste sera explorée à travers un travail d'équipe avec des collègues de LASTIG et sera présenté en Section 4.4

4.2 Identification des mises à jour d'un référentiel du réseau routier d'autorité en utilisant les traces issues des activités sportives

Avant-propos

Cette section présente la deuxième partie des travaux de la thèse de Stefan Ivanovic soutenue en 2018 qui est focalisée sur la mise à jour de données topographiques d'autorité à partir de traces issues des activités sportives en zone de montagne.

Cette section résume donc les travaux de la thèse et d'un article de conférence :

Ivanovic, Stefan S., Ana-Maria Olteanu-Raimond, Sébastien Mustière, and Thomas Devogele. 2019. 'Potential of Crowdsourced Traces for Detecting Updates in Authoritative Geographic Data'. In *Geospatial Technologies for Local and Regional Development*, edited by Phaedon Kyriakidis, Diofantos Hadjimitsis, Dimitrios Skarlatos, and Ali Mansourian, 205–21. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing.

4.2.1 Méthode pour l'identification d'alertes de mise à jour du référentiel routier

L'hypothèse explorée par S. Ivanovic a été que, malgré les limites des traces provenant des activités sportives, elles sont une source riche pour la mise à jour de données d'autorité. L'approche proposée est composée de deux étapes. La première consiste à appairer les traces GNSS avec les données d'autorité. La deuxième consiste à définir si un segment ou un ensemble de segments non appariés doit être considéré comme une alerte pour la mise à jour.

4.2.1.1 Appariement des traces GNSS

Dans ces travaux de recherche, l'appariement de données a été utilisé comme outil pour l'identification de mises à jour potentielles. L'appariement consiste à identifier des objets géographiques homologues entre deux ou plusieurs sources de données géographiques différentes qui représentent la même réalité du monde réel. Les résultats d'appariement sont représentés sous forme de liens d'appariement ayant plusieurs cardinalités : 1:0, 1:1, n:m, ou 0:1. Dans notre contexte, la cardinalité qui nous intéresse est 1:0 (*i.e.* un objet géographique d'une source de données n'a pas d'homologue dans une autre source de données). L'hypothèse est que la source de données ne contenant pas les objets géographiques non-appariés pourrait faire l'objet d'une mise à jour et que les objets géographiques non-appariés peuvent être considérés comme candidats, soit pour être intégrés dans le référentiel d'autorité, soit pour créer une alerte de mise à jour.

Comme déjà évoqué en Section 3.3, le manque de protocole, la faible exhaustivité et l'incomplétude des métadonnées et de l'information thématique ont rendu difficile l'usage des approches d'appariement proposées dans la littérature et qui sont basées sur des raisonnements plus complexes (Mustière et Devogele, 2008; Newson et Krumm, 2009; Olteanu-Raimond et al., 2015). Par conséquent, l'approche d'appariement proposée est une approche géométrique basée sur la notion de buffer grandissant inspirée des travaux de (Walter et Fritsch 1999; Liu et al. 2015).

La méthode proposée peut être résumée comme suit. Pour chaque segment d'une trace GNSS (notée DS_1 en Figure 26), l'algorithme cherche des segments candidats à l'appariement dans le référentiel d'autorité (notée DS_2 en Figure 26). Cette étape, nommée la sélection des candidats, s'appuie sur la notion de buffer grandissant et utilise l'exactitude des points GNSS décrite en Section 3.3. Ainsi, la taille du buffer varie en fonction de l'exactitude des segments de la trace qui a été estimée à partir des exactitudes des deux points composant chaque segment. La taille du buffer est plus grande pour des segments de faible exactitude et plus petite pour les segments de bonne exactitude (Figure 26).

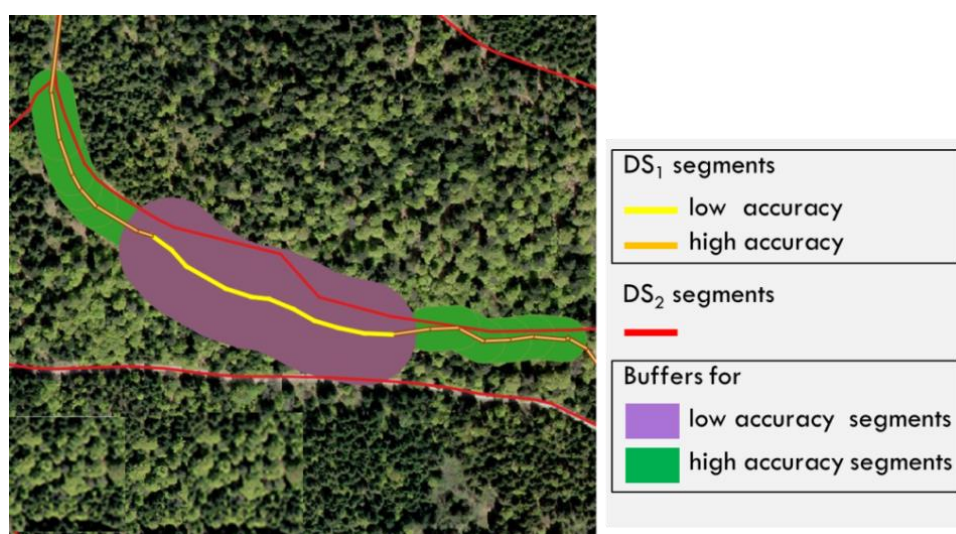


Figure 26. Sélection des candidats à l'appariement en fonction de l'exactitude des segments des traces où DS_1 – (DataSet₁) signifie les traces issues d'activités sportives et DS_2 – (DataSet₂) représente des objets géographiques provenant du réseau routier d'autorité de la BDTOPO®; issue de (Ivanovic et al. 2019)

Ensuite, chaque couple (segment de trace, candidat) est analysé en utilisant trois critères, à savoir un critère de direction (*i.e.* deux segments colinéaires ont plus de chance d'être homologues que deux segments perpendiculaires), un critère de voisinage (*i.e.* un segment a plus de chance d'être apparié si ses voisins sont appariés), et un critère de longueur (*i.e.* pour une mise à jour, le segment à intégrer dans le référentiel d'autorité devrait avoir une longueur minimale, afin de respecter les spécifications des données, et d'éviter une mise à jour insignifiante considérée comme un bruit).

Les critères sont appliqués l'un après l'autre, d'abord le critère de direction, puis le critère de voisinage ; enfin, après avoir agrégé les segments consécutifs appariés et non-appariés, le critère de longueur a été appliqué. Seuls les segments agrégés non-appariés dont la longueur est supérieure à un certain seuil sont considérés des potentiels candidats pour la mise à jour.

4.2.1.2 Identification des alertes de mise à jour

A la fin de l'étape d'appariement de données, les traces sont composées d'un ensemble de segments qualifiés selon deux étiquettes « apparié » ou « non-apparié ». Les segments non-appariés peuvent faire l'objet d'une alerte pour la mise à jour du réseau routier d'autorité. Cependant, déterminer quel segment non-apparié représente une véritable mise à jour, est une tâche difficile. Différentes interprétations sont possibles pour les segments non-appariés, comme l'illustre par exemple la Figure 27 : (1) la déviation d'une partie de la trace par rapport à la route faisant autorité est causée par des mesures d'erreur GPS; (2) la déviation suit une route existante dans le monde réel qui n'est pas représentée dans les données d'autorité ; (3) la déviation est due à une modification de la route existante. Les deux derniers cas sont considérés comme des routes manquantes, donc des mises à jour potentielles et différentes.

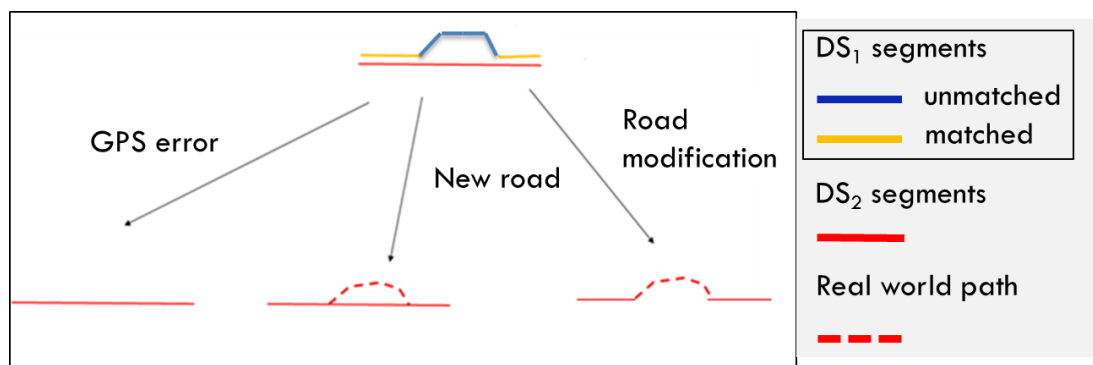


Figure 27. Différentes interprétations pour un segment non-apparié : (1) erreur GPS, (2) nouvelle route, (3) modification de la route ; issue de (Ivanovic et al., 2019)

Etant donné les différentes interprétations que l'on peut avoir, l'objectif est de qualifier les segments non-appariés afin de faciliter le processus de mise à jour. La qualification est faite en s'appuyant sur un coefficient de confiance. Ce dernier est calculé en se basant sur quatre critères. Un premier critère est basé sur la redondance des traces. L'hypothèse est que plus il y a de segments de traces non-appariées suivant les mêmes itinéraires, plus on est confiant qu'il existe une route sur le terrain. Un deuxième critère porte sur l'actualité des traces et modélise la fraîcheur de l'information. Le troisième critère s'appuie sur l'exactitude des segments déterminée selon la méthode décrite en Section 3.3.3. L'hypothèse est qu'un segment non-apparié ayant une bonne exactitude est plus fiable qu'un segment non-apparié ayant une faible exactitude. Enfin, le quatrième critère mesure la continuité avec laquelle une trace suit un sentier du référentiel d'autorité. En effet, il se peut qu'une trace suive un sentier du réseau routier d'autorité sur une longue portion de sa longueur, sauf dans certaines parties. Certains de ces cas ne doivent pas être considérés comme des mises à jour, et peuvent être dus aux hétérogénéités concernant la qualité géométrique des traces ou encore au comportement humain (*e.g.* le randonneur peut quitter le sentier pour raccourcir le trajet). D'autres, au contraire, doivent être considérées comme des mises à jour, car la trace suit réellement un

chemin qui n'est pas représenté dans les données faisant autorité. Dans ce cas de figure, le critère basé sur la continuité permet de distinguer les segments non-appariés suivant le même sentier de ceux qui ne suivent pas le même sentier.

Chaque critère associe un indicateur de confiance inclus dans l'intervalle [0-1], où 0 signifie que la confiance dans l'hypothèse selon laquelle le segment non-apparié doit générer une alerte de mise à jour est très faible, tandis que 1 signifie que la confiance attribuée à la même hypothèse de mise à jour est forte. Afin de calculer la confiance totale, une méthode classique de combinaison basée sur l'égalité de poids est utilisée (Roszkowska, 2013).

Pour plus d'information sur les règles définies pour les critères ainsi que la combinaison des critères, j'invite le lecteur à se référer à (Ivanovic et al. 2019).

4.2.2 Application et résultats

Les deux approches présentées ci-dessus ont été appliquées pour identifier des mises à jour concernant les sentiers de la BDTOPO® de l'IGN à partir des traces de randonnées situées dans le massif des Vosges (voir la Figure 23).

Les résultats d'appariement ont été validés manuellement pour un échantillon de 41 traces (920 km) par trois experts en se basant sur différentes sources de données telles que des cartes ou orthophotos. Une précision et un rappel de respectivement 0.84 et 0.77 ont été obtenus. Ces résultats nous semblent équilibrés, étant donné le contexte d'aide à la décision pour un opérateur travaillant sur la mise à jour de données d'autorité.

En ce qui concerne les alertes de mise à jour, un total de 727.1 km de sentiers manquants ont été identifiés, ce qui correspond à 7.7% de la longueur totale des traces GNSS. Comparé à la longueur totale des sentiers de la BDTOPO®, cela représente une contribution de 2.4%. En ce qui concerne le degré de confiance, 19% des routes manquantes ont une confiance faible (confiance totale < 0.4), 66% ont une confiance moyenne (confiance comprise entre 0.4 et 0.6) et 15 % ont une confiance forte (confiance totale > 0.6).

L'analyse visuelle des résultats a permis de différencier trois configurations typiques de sentiers manquants qui peuvent lever des alertes de mise à jour (voir la Figure 29).

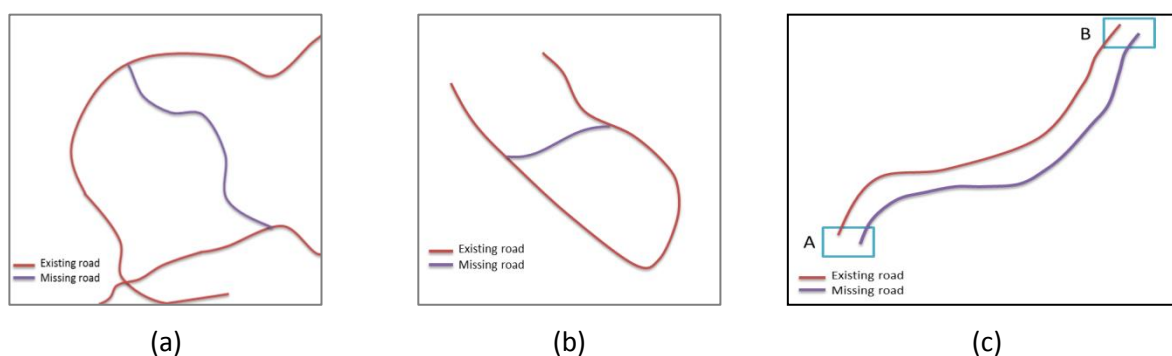


Figure 28. Configurations typiques de sentiers d'autorité manquants détectés : (a) sentiers longs indépendants; b) raccourcis; (c) sentiers parallèles aux sentiers d'autorité ; issue de (Ivanovic et al., 2019)

La première configuration, de type sentier long indépendant (Figure 29 a), se caractérise par une longueur importante et permet un itinéraire différent. Ce type de sentier représente 51,8 km soit 42% du total des sentiers détectés. Le deuxième type est le raccourci, comme l'illustre la Figure 29 (b). Ils ont un usage complémentaire permettant généralement de gagner du temps ou de réduire la distance entre une origine et une destination. Sur le nombre total de sentiers manquants, les raccourcis représentent 26 % (12,1 km). Enfin, la dernière catégorie, appelée sentier parallèle (Figure 29 c), est caractérisée par une partie de la trace parallèle à un sentier du réseau d'autorité sur une

longue distance, ayant le même itinéraire qu'un chemin autorisé, par exemple reliant les mêmes origines et destinations. Les sentiers parallèles représentent 22,4 km soit 32% des sentiers détectés comme des alertes potentielles de mise à jour.

La Figure 29 illustre des exemples des sentiers manquants détectés dans la zone de test selon ces cas de figure.

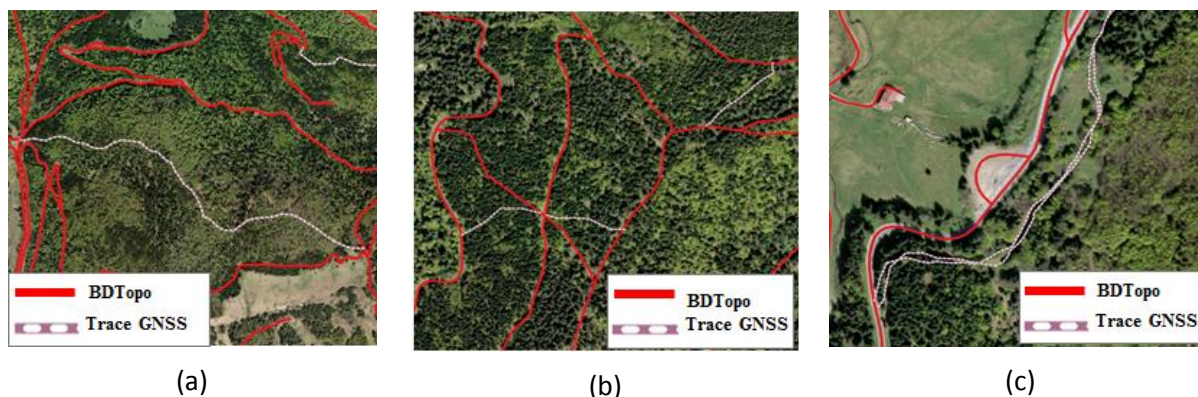


Figure 29. Exemples de sentiers manquants détectés à partir des traces GNSS: (a) sentiers longs indépendants; (b) sentiers identifiés comme potentiels raccourcis; (c) sentiers parallèles aux sentiers de référence ; issue de (Ivanovic et al., 2019)

Dans l'ensemble, les sentiers manquants les plus représentés sont les sentiers longs indépendants, qui ont la plus haute priorité pour les ANC. Avec un peu plus d'un quart, les raccourcis ne sont pas négligeables dans les chemins manquants. Cela s'explique principalement par le comportement humain, c'est-à-dire le gain de temps et de distance dans leur mobilité en prenant généralement des raccourcis lorsque cela est possible, surtout en zone ouverte comme en zone de montagne. Les routes parallèles, qui représentent près d'un tiers des chemins manquants, présentent moins d'intérêt pour les ANC car ils suivent le même itinéraire que les chemins d'autorité existants. Ainsi, selon les spécifications de la base de données, il n'est pas toujours nécessaire de les prendre en considération.

4.3 Mise à jour de données d'occupation du sol d'autorité

Avant-propos

Cette section présente les travaux du post-doctorat réalisé par Lanfa Liu que j'ai co-encadré avec mes collègues Laurence Jolivet et Arnaud Le Bris. Ce travail a été réalisé dans le cadre du projet LandSense.

Cette section résume donc un livrable soumis et validé par la commission européenne et qui a fait l'objet d'une soumission en cours d'évaluation :

Liu, L., Olteanu-Raimond, A.-M., Jolivet, L., Le Bris, A., 2019. Operational workflows for integration of citizen-observed data in authoritative systems (Livrable No. D5.5).

4.3.1 Fusion d'informations géographiques volontaires pour la mise à jour de données d'occupation du sol

L'objectif de ce travail a été d'étudier le potentiel de l'IGV pour la mise à jour des données d'occupation du sol. Un des enjeux a été d'aller plus loin que la remontée d'alertes dans le processus de mise à jour. Nous avons identifié deux types de mise à jour : la **mise à jour** engendrée par les changements dans le monde réel et la **mise à niveau** due à l'enrichissement thématique. Cette dernière consiste à proposer une classification plus fine de données d'occupation du sol.

Ce travail s'inscrit à la suite des travaux présentés en Section 2.3.2 et 3.2.2 et décrit la fusion de plusieurs sources d'IGV collectées grâce aux outils développés dans le projet LandSense et aux campagnes de collecte organisées.

Notre méthode s'appuie sur l'hypothèse selon laquelle pour un millésime donné d'une base de données d'occupation du sol (OCS), nous sommes en mesure de générer des alertes de mise à jour provenant de différentes sources d'information et portant sur les deux aspects thématique et géométrique. Chaque objet géographique de la base de données d'OCS peut faire l'objet d'une ou plusieurs alertes. Certaines peuvent générer une mise à jour (par exemple un changement de classe) ou une mise à niveau (classification avec une classe de niveau inférieur), d'autres, en revanche, peuvent ne pas avoir d'impact (*e.g.* de nouvelles constructions dans une zone qui est déjà une zone bâtie). Cependant, étant donné l'hétérogénéité des sources d'information et le manque d'exhaustivité, si une zone ne contient pas d'alerte, nous ne pouvons pas émettre l'hypothèse que la zone n'a pas changé. La Figure 30 illustre un espace délimité par cinq polygones de classes d'usage différentes. Le polygone E contient 3 alertes provenant de trois sources différentes, le polygone B contient 2 alertes provenant d'une seule source, le polygone A ne contient aucune alerte, etc.

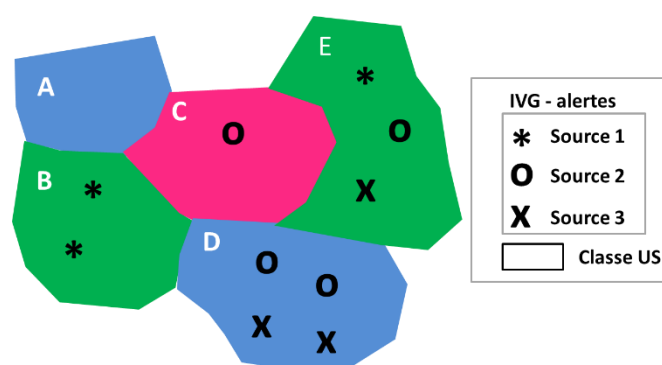


Figure 30. Illustration de cinq polygones d'OCS caractérisés par zéro (A), un (C), deux (B), trois (E) et quatre (D) alertes provenant de trois sources d'IGV.

Afin de prendre en compte l'hétérogénéité et parfois le manque de qualité des sources d'IGV, nous proposons une approche basée sur la fusion d'informations. Pour cela, nous nous appuyons sur la théorie des fonctions de croyance qui s'est avérée adaptée à la fusion de sources d'information géographiques et qui permet de prendre en compte les imprécisions, incomplétudes ou incertitudes (Olteanu-Raimond et al., 2015). Dans ce travail de recherche, nous nous sommes focalisés sur la mise à jour de classes d'usage de l'OCS et nous avons utilisé quatre sources d'IGV qui seront décrites en Section 4.3.2. Notons que pour ce premier pas vers la mise à jour automatique, la mise à jour des limites de classes n'a pas fait l'objet de nos recherches.

L'approche proposée est composée de quatre étapes principales.

La première étape est le prétraitement. Elle consiste à associer à chaque objet géographique de la base de données d'OCS l'ensemble des alertes venant des sources d'IGV à notre disposition. Pour les alertes représentées par des objets géographiques sous forme de polygones, un processus de découpage géométrique a lieu en utilisant la couche de l'OCS. Ainsi, une alerte appartient à un seul et unique objet géographique d'OCS. Afin d'éliminer les objets parasites créés par le découpage, un filtre basé sur l'aire des polygones est appliqué. Seuls les morceaux d'alertes ayant la valeur de l'aire supérieure à un seuil sont gardés.

La deuxième étape consiste à appairer les nomenclatures des sources d'IGV avec la nomenclature de l'OCS d'autorité. Cette étape, nécessitant une connaissance approfondie de la nomenclature de l'OCS, a été réalisée manuellement dans le cadre du post-doc de L. Liu. Cependant, des outils d'appariement de schémas ou de l'alignement de nomenclatures tels que EAGLE (Arnold et al., 2016) peuvent être utilisés.

La troisième étape représente l'étape de fusion d'informations basée sur la théorie de fonctions de croyance. Je ne détaille pas ici le cadre théorique de la théorie de fonctions de croyance ; le lecteur est invité à voir la publication (Olteanu-Raimond, et. al 2015).

Cette étape de fusion nécessite tout d'abord la définition du cadre du discernement et l'initialisation des masses de croyance. Ainsi, pour chaque objet de l'OCS qui possède au moins une alerte, un cadre de discernement est composé d'hypothèses simples, chaque hypothèse représentant une classe d'usage (e.g. LU2 - Usage Industriel). Le cadre de discernement est défini par l'équation 1 :

$$\Theta = \{LU1.1, LU1.2, \dots, LU2, LU3, LU4.1.1, \dots, LU5, LU6.1, \dots, LU6.3\} \quad (1)$$

A partir du cadre de discernement, nous déterminons l'ensemble de toutes les combinaisons possibles :

$$2^\Theta = \{\{LU1.1\}, \{LU1.2\}, \dots, \{LUi, LUj\}, \dots, \Theta, \phi\} \quad (2)$$

où $\{\{LUi, LUj\}\}$ est un sous-ensemble de la proposition P, qui peut être interprété comme l'hypothèse que la classe d'usage de l'objet géographique est soit LUi ou LUj et ϕ est l'ensemble vide. Une fois toutes les combinaisons déterminées, une masse de croyance représentant la croyance avec laquelle la source d'IGV croit dans la véracité d'une proposition P est calculée avec la condition que la somme de toutes les masses de croyance est égale à 1 et la masse de croyance attribuée à l'ensemble vide est égale à zéro.

Dans notre approche, nous avons proposé deux types d'initialisation des masses de croyance : pour les sources dont l'alerte est un polygone (m_{ipoly}) et pour les sources ponctuelles (m_{jpoint}).

Ainsi, pour chaque polygone de l'OCS affecté par au moins une alerte,

$$m_{ipoly}(P) = AC * \frac{Area(alerte\ polygone)}{Area(polygone\ OCS)}, i=1..N, P \subseteq 2^\Theta \quad (3)$$

où $m_{ipoly}(P)$ représente la masse de croyance attribuée à la proposition P par la source i contenant des alertes de type polygone et AC représente l'accord de contributeurs tel que décrit dans la Section 3.2.2.2.

Etant donné que dans les campagnes que nous avons menées, une alerte ponctuelle fait référence à la totalité de l'objet géographique d'OCS, seul l'accord des contributeurs est pris en compte pour l'initialisation des masses de croyance pour les sources d'IGV contenant des alertes ponctuelles. Ainsi, la masse de croyance attribuée à la proposition P est définie comme dans l'équation 4.

$$m_{jpoint}(P) = AC, j=1..K, P \subseteq 2^\Theta \quad (4)$$

Afin de prendre en compte le fait qu'une source d'information n'a pas assez de connaissances pour exprimer sa croyance dans une proposition du 2^Θ , nous attribuons une masse de croyance à l'ignorance (i.e. l'ensemble d'hypothèses simples du cadre de discernement). La masse de croyance attribuée à l'ignorance est calculée par les équations (5) et (6).

$$m_{ipoly}(\Theta) = 1 - \sum_{P \subseteq 2^\Theta} m_{ipoly}(P) \quad (5)$$

$$m_{jpoint}(\Theta) = 1 - \sum_{P \subseteq 2^\Theta} m_{jpoint}(P) \quad (6)$$

Les masses de croyance obtenues après cette étape d'initialisation sont fusionnées en utilisant l'opérateur de Dempster. Il s'agit d'une règle conjonctive déterminant une nouvelle masse de croyance exprimant la combinaison de croyances provenant de toutes les sources. Cette règle conjonctive peut générer du conflit. Par exemple, si la Source 1 croit, avec une croyance de 0.5, que la classe de l'objet géographique est LU3 (Commercial), et la source 2 croit, avec une croyance de 0.7, que la classe du même objet géographique est LU5 (Résidentiel), la fusion des deux sources va

généraliser un conflit ayant une masse de croyance de 0.35. L'opérateur de Dempster propose de redistribuer le conflit sur l'ensemble de masses de croyance non-nulles issues de la fusion.

Le dernier concept utilisé du cadre de la théorie des fonctions de croyance est la probabilité pignistique. Elle permet de déterminer les masses de croyance attribuées aux hypothèses singletons tout en prenant en compte la part des masses de croyance des hypothèses composées qui les contiennent.

$$\forall Pr \in \Theta, Pr(P) = \sum_{P \subseteq P', P' \subseteq \Theta} m(P') \frac{|P \cap P'|}{|P'|} \quad (7)$$

Enfin la quatrième étape permet de prendre la décision finale. En s'appuyant sur le maximum de la probabilité pignistique, trois décisions sont possibles : (1) mise à niveau (enrichissement), (2) mise à jour (changement) et (3) pas de mise à jour. Pour les deux premiers cas, la méthode est en mesure de fournir la nouvelle classe d'usage. Chaque décision est qualifiée par un indicateur de confiance mesuré par la longueur de l'intervalle [crédibilité-plausibilité].

4.3.2 Applications et résultats

L'approche proposée par Lanfa Liu a été appliquée sur des données réelles. D'une part le millésime 2016 de l'OCS-GE® de l'IGN, d'autre part les données issues des campagnes de collecte que nous avons effectuées sur notre zone d'étude (voir la Figure 31) en suivant les protocoles définis et en utilisant les outils Paysages développés (web et application mobile). Il convient de mentionner que grâce aux protocoles de collecte, les IGV collectées sont millésimées 2019, ce qui nous permet d'obtenir un millésime 2019 pour l'OCS-GE.

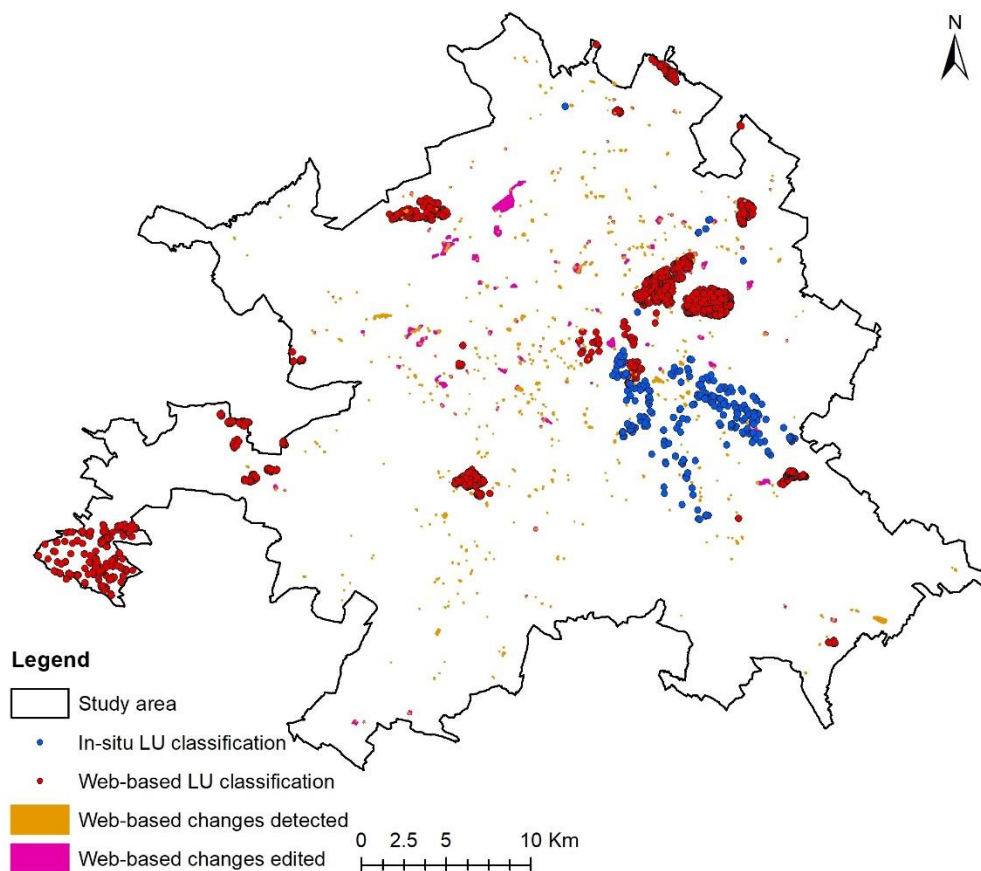


Figure 31. Sources d'informations géographiques volontaires dans la zone d'étude ; issue de (Liu et al., 2019)

Le Tableau 8 synthétise les données collectées selon les sources :

- Source 1 (Web-based changes detected) : les alertes représentant les changements détectés automatiquement par le service de changement et validés par différents profils d'utilisateurs avec l'application web Paysages ;
- Source 2 (Web-based changes edited) : les alertes représentant les changements édités par les contributeurs avec l'application Web Paysages;
- Source 3 (Web-based LU classification) : alertes représentant la classification d'usage de certaines zones bâties réalisée par les contributeurs en utilisant l'application web Paysages.
- Source 4 (In-situ LU classification) : alertes représentant la classification d'usage réalisée par les contributeurs sur le terrain avec l'application mobile Paysages ;

Tableau 8. Synthèse de données utilisées pour la mise à jour de données d'autorité d'OCS-GE ; adapté depuis (Liu et al., 2019)

	Source 1 : Web-based changes detected	Source 2: Web-based changes edited	Source 3: Web-based LU classification	Source 4 : (<i>In-situ</i> LU classification)
Format de l'alerte	Polygone	Polygone	Point (centroïde de l'objet OCS-GE)	Point (centroïde de l'objet OCS-GE)
Nombre d'alertes	650	142	2711	304
Nombre d'observations total (une alerte peut avoir plusieurs observations)	2778	142	2711	369
Nombre d'observations par objet OCS-GE	Entre 3 et 6	1	1	Entre 1 et 4
Nombre possibles de classes d'usages distinctes	7	7	7	17
Types de classes d'usages collectées	Résidentiel; Industriel; Infrastructure; En construction; Autre; Pas de changement; Inconnu	Résidentiel; Industriel; Infrastructure; En construction; Autre; Pas de changement; Inconnu	Résidentiel; Industriel; Commercial; Industriel ou Commercial; Industriel, Commercial ou Résidentiel; Autre; Inconnu	Activités d'extraction ; Zones en construction Agriculture; résidentiel; Industriel ; Infrastructure

Le workflow décrit ci-dessus a été appliqué sur notre zone d'étude comportant 171 621 objets géographiques OCS-GE. Les résultats suivants ont été obtenus : 25 objets OCS-GE (0,1 km²) sont classés en « pas de mise à jour », 409 objets OCS-GE (7,36 km²) sont classés en « mise à jour (changement) » et 2975 objets OCS-GE (5,75 km²) en « mise à niveau (enrichissement) ». Environ 87,3 % des polygones détectés ont été re-classifiés en classes d'usage d'un niveau hiérarchique plus fin, passant ainsi d'une classification US235 (Usage Industriel, Commercial et Service et Résidentiel) en US2 (Usage Industriel), US3 (Usage Commercial et Service) et US5 (Usage Résidentiel).

Quelques résultats sont illustrés en Figure 32. Les alertes issues des changements détectés automatiquement (Source 1) et celles issues des changements édités (Source 2) sont représentées respectivement en orange et en rose. Les points bleus représentent les lieux visités sur le terrain pour la classification (Source 4) tandis que les points rouges représentent les lieux qui ont fait l'objet d'une classification des zones bâties (Source 3). Les images Pléiades de 2019 et Orthophoto de 2016 du Géoportail français ont été utilisées pour la comparaison visuelle. La Figure 32-a illustre un exemple où un objet OCS-GE de classe US6.3 (Sans usage) en 2016 contient quatre alertes de changements détectés : une provenant de la Source 2 (US 5 – Résidentiel) et trois de la Source 1 (deux de classe US5-Résidentiel ; une de classe US4.1-Réseau de transport). La décision choisie par notre méthode est « Mise à niveau (enrichissement) ». La classe d'usage de l'objet OCS-GE passe donc d'un usage LU6.3 (Sans usage) à un usage US5 (Résidentiel). Pour le second exemple (voir Figure 32-b), l'objet OCS-GE de classe US4.1.4 (Transport de l'eau) est caractérisé par une alerte *in-situ* (Source 4) et un changement détecté (Source 1). La décision finale est « Pas de mise à jour », la classe d'usage reste LU4.1.4. Dans la Figure 32-c, l'objet OCS-GE de classe US235 (Industriel, Commercial et Services, et Résidentiel) possède une alerte issue de la Source 3 de type US5 (Résidentiel). La décision prise est « Mise à niveau (enrichissement) » et l'objet OCS-GE classé en US235 passe en US5 pour le millésime 2019.

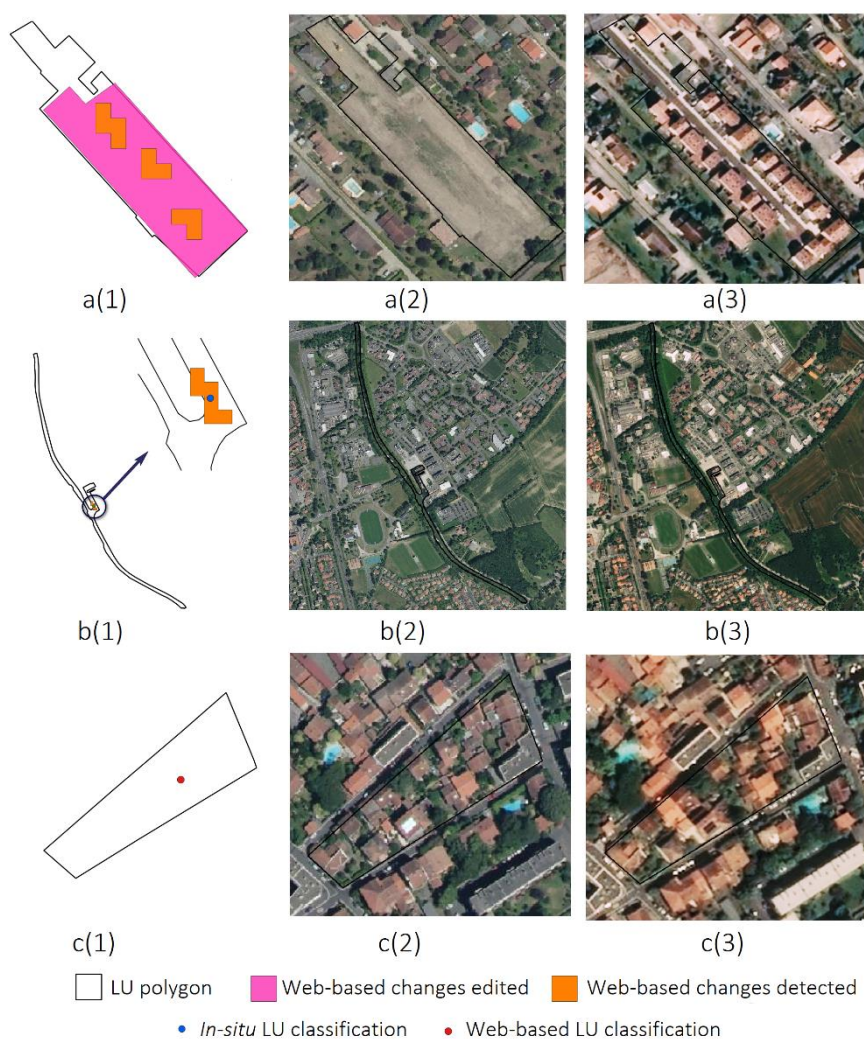


Figure 32. Exemples de polygones OCS mis à jour grâce aux sources d'IGV collectées. (a1-c1) polygones US et sources d'IGV associées ; (a2-c2) Orthophoto (©IGN) en 2016 ; (a3-c3) Images Pléiades en 2019 ; issue de (Liu et al., 2019)

Afin d'évaluer les résultats, 144 objets géographiques d'OCS-GE qui ont fait l'objet d'une décision ont été manuellement validés en comparant les images de 2016 et 2019. Pour les cas dont l'estimation de l'usage a été difficile, l'imagerie immersive de Google Street View a été également utilisée. Une précision de 0.78 a été obtenue. Nous avons étudié la relation entre la précision globale et la mesure de confiance calculée pour chaque décision en utilisant l'intervalle de crédibilité et plausibilité. Le seuil de confiance a d'abord été fixé à 0, ce qui signifie que tous les objets OCS mis à jour sont justes. Lorsque l'on augmente le seuil de confiance, la précision commence à s'améliorer, puis diminue légèrement. La précision la plus élevée (0.93) est obtenue lorsque le seuil de confiance est fixé à 0.5, ce qui correspond à 46 objets (31.9%) mis à jour parmi l'échantillon de 144.

Les résultats montrent le potentiel de l'IGV pour la mise à jour de données d'occupation du sol. L'approche proposée présente de nombreux avantages tels que la gestion multi-dates pour mettre à jour des données millésimées, la fusion de sources d'IGV hétérogènes spatialement, thématiquement, incomplètes et de qualité différente, ainsi que le potentiel d'évolution et d'adaptation de l'approche. En effet, plusieurs sources d'information peuvent être ajoutées, volontaires ou d'autorité, ou l'approche peut être adaptée pour mettre à jour d'autres données d'OCS d'autorité tels que CorineLandCover. Notre approche présente également des limites. Premièrement, la mise à jour ne concerne pas les changements de limite des objets OCS-GE. Deuxièmement, les sources d'IGV ne présentent que peu d'information redondante, ce qui pose problème pour l'initialisation des masses de croyance, qui nécessiterait une analyse de la sensibilité. Enfin, la validation complète de nos résultats n'a été réalisée que sur un échantillon de 144 objets, que nous considérons comme faible. Pour obtenir une validation avec un niveau de confiance de 95% et une marge d'erreur à 5%, l'échantillon de validation devrait être composé de 384 objets (Foody, 2009).

4.4 Constitution d'un référentiel métier pour la localisation des victimes en montagne

Avant-propos

Cette section présente les travaux initiés avec mes collègues de LASTIG impliqués dans le projet Choucas ayant comme objectif de proposer une méthode pour la constitution d'un référentiel métier des objets de repère et d'itinéraires en zone de montagne.

Je résume ici les travaux faisant l'objet d'un livrable portant sur la définition d'une ontologie des objets de repères (OOR) et d'une publication analysant le potentiel des sources d'IGV pour l'aide à la localisation des victimes en montagne (Van Damme et al., 2019).

Olteanu-Raimond, A-M., Mustière, S., Bunel, M., Domingues, C., Duchêne, C., Jolivet, L., et Van Damme, M-D. 2020. 'Ontologies d'objets de repères Choucas'.

Van Damme, M-D., Olteanu-Raimond, A-M and Méneroux, Y, 2019. 'Potential of Crowdsourced Data for Integrating Landmarks and Routes for Rescue in Mountain Areas'. *International Journal of Cartography* 5 (2–3): 195–213.

Comme évoqué en Section 4.1, la constitution d'un référentiel métier est un besoin nécessaire pour pouvoir localiser les victimes en zone de montagne à partir des indices donnés par les requérants. Afin de constituer un référentiel métier, une première étape a été d'identifier les éléments qui peuvent servir comme référence pour décrire leur localisation (itinéraire, lac, cabane, sommet) et de définir la manière dont ces éléments peuvent être combinés afin d'améliorer le processus de localisation (e.g. identifier tous les lacs qui sont proches d'un itinéraire). La deuxième étape a été de générer des connaissances sur la disponibilité, le format et la complémentarité des sources d'information d'autorité et volontaires.

4.4.1 Modélisation des objets de référence

Dans les travaux de thèse de Mattia Bunel publiés dans Bunel et al. (2019), un indice du type : « Je suis presque au sommet du Grand Mont » est modélisé sous la forme d'un triplet composé d'un site (« je »), d'une cible (sommet du Grand Mont) et d'une relation de localisation étendue indiquant la position du site par rapport à la cible, conformément à l'équation 8.

$$indice = < site, relation\ de\ localisation\ étendue, cible > \quad (8)$$

La relation de localisation étendue, quant à elle, est définie par trois éléments, à savoir le verbe, la relation de localisation (être) et le modifieur (presque). La composition de la relation de localisation étendue varie en fonction de l'indice, le verbe et la relation de localisation étant toujours présents tandis que le modifieur est un élément facultatif.

$$relation\ de\ localisation\ étendue = < verbe, modifieur, relation\ de\ localisation > \quad (9)$$

La modélisation de l'indice permet de passer d'un référentiel spatial indirect à un référentiel spatial direct, c'est-à-dire de construire une zone de localisation compatible (Viry et Villanova-Oliver 2019) pour chaque indice. La méthodologie de construction d'une zone de localisation compatible avec l'indice implique trois objectifs : (1) construire un référentiel métier contenant les types d'objet de référence, (2) identifier les types de relation de localisation utilisés dans un espace ouvert, tel que la montagne, et (3) proposer des méthodes de spatialisation. Ces dernières dépendent de la relation de localisation étendue et du type d'objet de référence.

Dans le cadre de ce manuscrit, j'ai fait le choix de ne décrire que les travaux de recherche entamés pour répondre au premier objectif. Le deuxième et le troisième objectifs ont fait l'objet des travaux de recherche réalisés dans le cadre de la thèse de Mattia Bunel.

La modélisation des indices s'appuie sur le formalisme OWL (Web Ontology Language) qui permet une plus grande interopérabilité et qui facilite la définition des mécanismes de raisonnement et de recherche d'information. L'instanciation de l'ontologie et le raisonnement permettant de combiner les éléments de localisation s'appuient sur des techniques d'intégration de données.

Nous avons identifié trois grandes familles d'objets de référence qui peuvent guider la localisation des victimes en montagne : les objets de repère, les itinéraires et les objets cartographiques. Les trois familles d'objets de référence sont définies dans trois ontologies d'applications, à savoir l'Ontologie des Objets Repères¹⁸ (OOR), l'ontologie des itinéraires (OOI) et l'ontologie des objets cartographiques (OOC). Ces trois ontologies sont interconnectées d'une part avec l'Ontologie d'Alertes Choucas¹⁹ (OAC), qui est l'ontologie mère définie dans le cadre du projet Choucas par mes collègues du laboratoire Steamer de l'Université de Grenoble (Viry et Villanova-Oliver, 2019), et d'autre part avec l'ontologie des relations de localisation (ORL) permettant d'identifier et d'organiser les relations de localisation utilisées pour se repérer en montagne, et qui a été définie dans le cadre de la thèse de Mattia Bunel.

L'ontologie OOR décrit les types d'objets qui permettraient à une personne de se repérer en montagne (e.g. lac, rivière, sentier, cabane, forêt, glacier, sommet, parking). Cette ontologie est focalisée sur la description du domaine de la montagne en France métropolitaine, et plus particulièrement sur ce qui est visible sur le terrain. L'ontologie OOR a été définie à partir de l'ontologie Géonto réalisée dans le cadre du projet ANR GEONTO (Mustière et al., 2011a). Notre méthode de travail a consisté à réorganiser l'ontologie initiale et supprimer les concepts dont l'objectif sortait de notre cadre d'étude. Pour la ré-organisation des éléments du relief, nous nous sommes appuyés sur une ontologie spécialisée sur les formes du relief LandFormReference Ontology (Sinha et al., 2018). Ensuite, cette nouvelle organisation a été complétée par des concepts issus de

¹⁸ <http://purl.org/choucas.ign.fr/oor>

¹⁹ <http://lig-tdcge.imag.fr/steamer/oac/index.html>

trois sources. La première source vient des experts en secours en montagne du PGHM de Grenoble. Il s'agit d'une liste des termes spécifiques au secours en montagne. La deuxième source provient des alertes réelles enregistrées sous forme audio. Un premier travail d'écoute et de retranscription d'alertes a permis d'identifier une liste de types d'objets de référence (Bunel et al., 2019a). C'est d'ailleurs ce travail qui nous a permis de voir que les victimes utilisent, entre autres, les symboles de la carte pour se localiser (e.g. « Je suis en bas du trait rouge de la carte »). Enfin, la troisième source d'information provient de l'analyse de sources d'IGV (voir la Section 4.4.2).

L'organisation est principalement hiérarchique, même si certains concepts peuvent hériter de plusieurs concepts parents. En suivant la modélisation des concepts du thésaurus RAMEAU de la BNF, chaque concept est caractérisé par un label principal (prefLabel) et un ou plusieurs labels secondaires (altLabel), exprimés en français et en anglais. Le label prefLabel exprime le label préférentiel tandis que les labels altLabel permettent de décrire des termes spécifiques ou des spécificités locales. Des propriétés permettant de modéliser les spécificités des objets ou celles du secours en montagne ont été définies tels que : *estComposéDe*, *peutEtreColocaliséAvec* (pour modéliser des relations particulières entre des objets liés sur le terrain comme les névés et les glaciers), *peutEtreConfonduAvec* (pour modéliser les objets qui peuvent être confondus de par leur forme, comme les lignes électriques et les câbles transporteurs), *aPourForme*, *aPourCouleur*. Il convient de noter que ce travail de définition des propriétés est un travail en cours qui avance au fur à mesure de l'avancement du projet. L'ontologie OOR disponible sur le site <https://choucas.ign.fr> contient au total 458 classes, associées à 1641 labels, en anglais et français, et une profondeur maximale de 6 niveaux. Les deux premiers niveaux d'organisation sont illustrés dans la Figure 33.

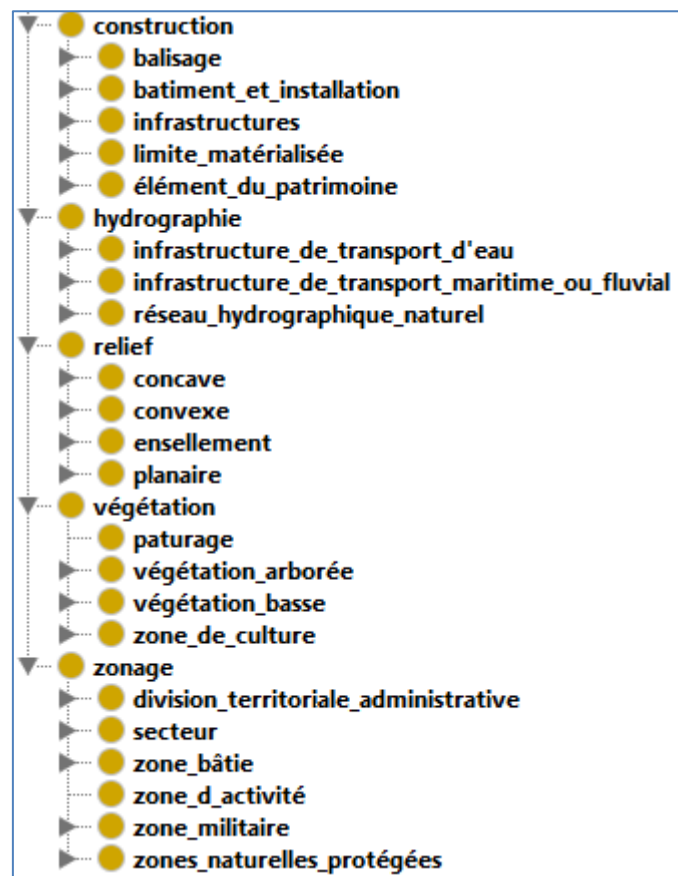


Figure 33. Premiers niveaux d'organisation de l'ontologie OOR ; issue de (Olteanu-Raimond et al., 2020 b)

Au premier niveau nous remarquons les concepts :

- construction : « regroupe toutes les constructions humaines dont la fonction apparente ne permet pas le rattachement aux classes hydrographie, végétation et relief »

- hydrographie : « contient tous les éléments hydrographiques et les constructions dont la fonction apparente est liée à l'hydrographie »
- relief : « les éléments du relief »
- végétation : « regroupe les éléments de végétation naturelle ou artificialisée »
- zonage : « regroupe les délimitations immatérielles du territoire, qui peuvent être de fait (agglomération) ou de jure (commune, zone industrielle) ». Définitions issues de (Olteanu-Raimond et al., 2020).

4.4.2 Enrichissement des sources d'autorité en utilisant les sources d'IGV

Afin d'identifier et évaluer le potentiel des données issues du web pour la constitution d'un référentiel métier, nous devons définir les relations sémantiques et les relations de correspondance entre les schémas des différentes sources d'information. Pour y parvenir, nous avons utilisé des techniques d'intégration telles que l'appariement de données ou le map-matching. L'hétérogénéité des données provenant de différentes sources rend la tâche d'appariement très complexe et surtout l'appariement des itinéraires nécessite une étape de pre-traitement plus importante.

La méthode proposée est composée de trois étapes. La première consiste à extraire les données issues de différentes sources d'IGV. Malgré le fait que l'accès aux données pour la plupart des sources se fait via des API, ce travail d'ingénierie est coûteux en temps, car les sources de données sont hétérogènes, ayant des schémas de données et de points de vue différents. Pour faciliter cette tâche, un plug-in QGIS²⁰ a été développé qui peut évoluer en fonction des besoins. La deuxième étape est l'appariement de données multi-sources (illustré en Figure 34) qui consiste à identifier les objets de repère homologues (*i.e.* deux objets de repère sont homologues s'ils représentent le même objet du monde réel) et les itinéraires homologues (*i.e.* deux itinéraires sont homologues s'ils suivent des chemins proches). La troisième étape, quant à elle, vise à analyser les caractéristiques des objets homologues afin d'identifier de nouveaux liens entre les différentes sources d'information et d'évaluer leur complémentarité. Pour cela, nous avons calculé des matrices de confusion d'une part, et des mesures de connectivité dans le réseau d'itinéraires d'autre part.

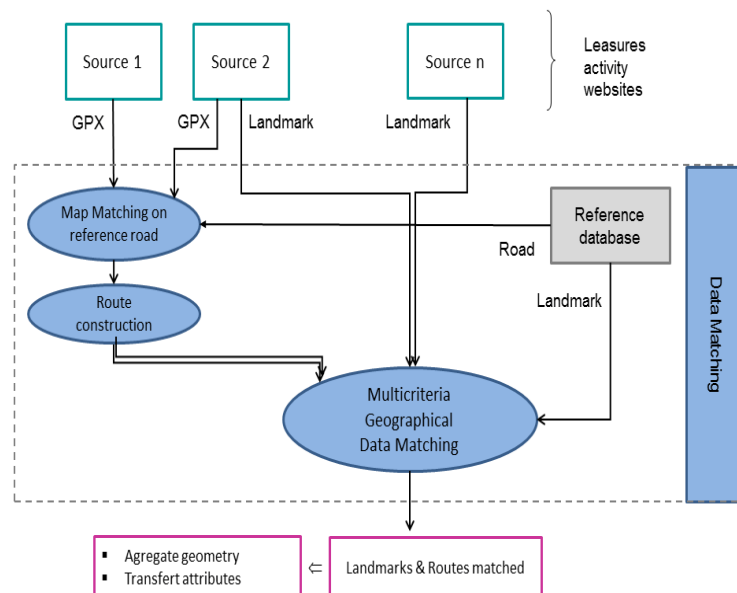


Figure 34. Approche proposée pour appairer les itinéraires et les objets de repère issus de sources d'IGV ; issue de (Van Damme et al., 2019)

²⁰ <https://github.com/ANRChoucas/PluginChoucas>

Je m'attarde davantage sur l'étape d'appariement de données multi-sources, que je considère comme primordiale dans le processus d'intégration. Comme évoqué en Section 3.3, les sites web mettant à disposition des traces fournissent rarement la géométrie linéaire des itinéraires, le plus courant étant la trace GNSS d'origine. Pour faciliter l'appariement des itinéraires, il est donc préférable de reconstruire leur géométrie. Deux types de pré-traitement sont proposés. Tout d'abord, les points de la trace vont être recalés sur un réseau routier de référence en appliquant l'algorithme de map-matching basé sur les modèles de Markov cachés (Newson et Krumm, 2009) qui permet de conserver la cohérence des itinéraires. Les résultats du recalage peut produire des tronçons non-connectés. Pour remédier à cet inconvénient, le deuxième pré-traitement consiste à reconstruire la géométrie linéaire de l'itinéraire en calculant le chemin le plus court sur le réseau routier entre 2 tronçons non connectés. Ainsi, à la fin du pré-traitement, la géométrie d'un itinéraire est composée d'une liste de tronçons du réseau routier. Ce travail a été réalisé par mes collègues Marie-Dominique Van Damme et Yann Meneroux.

L'appariement de données est effectué séparément pour les itinéraires et les objets de repère en utilisant l'algorithme d'appariement de données géographiques multi-critères qui est open-source²¹ et s'appuie sur la théorie des fonctions de croyance (Olteanu-Raimond et al., 2015). L'algorithme d'appariement nécessite de choisir une source de données de référence et une source de données de comparaison donnant ainsi une direction à l'appariement. Ainsi, pour chaque objet géographique de la source de référence, qui peut être d'autorité ou bien une source d'IGV, l'algorithme recherche des objets géographiques homologues dans le jeu de comparaison. Deux critères d'appariement (position et nom) ont été utilisés pour l'appariement des itinéraires et trois critères (position, nom et sémantique) pour l'appariement des objets de repère.

4.4.3 Application et résultats

La zone d'étude correspond à celle que nous avons fixée dans le projet Choucas, à savoir un territoire couvrant sept massifs montagneux dans les Alpes françaises, près de Grenoble (9 870 km²) (voir la Figure 35). Après l'analyse de différentes sources disponibles, nous avons focalisé notre étude sur celles dédiées aux sports de loisirs de montagne où il y a une communauté de contributeurs très active et qui sont recommandées par les autorités ou les pratiquants. Il s'agit de deux sources d'autorité, à savoir la BDTOPO® de l'IGN et les Aires Protégées (Geotrack) produites par les Parcs naturels régionaux ou les Parcs nationaux et trois sources d'IGV : Refuge.info, Camptocamp (C2C) et Visorando. D'autres sources d'IGV existent telles que OpenStreetMap (OSM), GeoNames, IGNRando, mais celles-ci n'ont pas été prises en compte, car elles ont déjà subi un processus d'intégration, la plupart des données étant importées massivement à partir d'autres sources.

Concernant les objets de repère, dans notre zone d'étude, les 25 906 objets sont répartis de la manière suivante : BDTOPO® (21097 objets), C2C (3229 objets), Refuge.info (478 objets), Geotrack (1102 points). Les 706 itinéraires se répartissent comme suit : Visorando (398), C2C (127) et Geotrack (181). Comme l'illustre la Figure 35, la couverture des points de repère et des itinéraires est bien diffuse. L'exactitude des objets de repère et des itinéraires est hétérogène d'une source de données à une autre. Le niveau de détail pour les itinéraires est également hétérogène au sein de la même source et entre les sources. En moyenne, la distance entre deux points consécutifs est de 25m, 50m et 5m pour Visorando, C2C et Geotrack respectivement. Pour plus d'information sur la description des sources, j'invite le lecteur à consulter l'article (Van Damme et al. 2019).

Etant donné nos objectifs (*i.e.* mesurer la complémentarité des sources d'IGV), trois appariements ont été réalisés pour les objets de repère : Camptocamp- BDTOPO®, Refuge.info- BDTOPO®, Geotrack- BDTOPO®, à chaque fois avec la BDTOPO® comme source de comparaison. En revanche,

²¹ <https://github.com/IGNF/MultiCriteriaMatching>

pour les itinéraires les appariements sont effectués deux à deux de la manière suivante : Geotrack - VisoRando ; Geotrack - Visorando et Camptocamp-Geotrack.

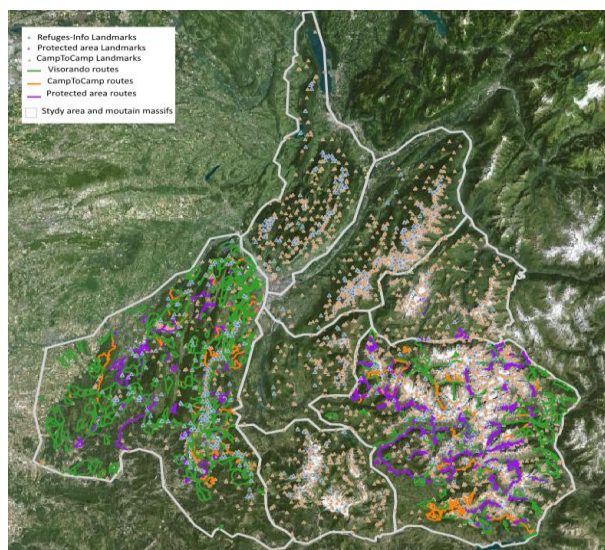


Figure 35. Zone d'étude pour la constitution d'un référentiel métier pour l'aide à la localisation des victimes en montagne ; issue de (Van Damme et al., 2019)

Les liens d'appariement de type 1 :1 (*i.e.* un objet d'une source de référence correspond à un objet d'une source de comparaison) obtenus grâce à l'algorithme d'appariement ont été validés manuellement. Les précisions obtenues sont de 0.95 pour Camptocamp-et Refuges.info et 0.71 pour Geotrack pour les objets de repère et de 0.85 pour les itinéraires. Pour les objets dont l'appariement est incertain, les liens d'appariement ont été corrigés et validés manuellement. Les résultats obtenus sont bons et fidèles à l'hétérogénéité des sources. Par exemple, les objets de repère et les itinéraires provenant des aires de protection (site Geotrack) sont très différents des autres du point de vue objectif, modèle, saisie et sémantique.

Complémentarité et redondance des sources

Le grand nombre d'objets non appariés (40% des itinéraires et 45% des objets de repère) prouve la richesse des sources d'IGV.

La distribution des objets de repère appartenant aux sources de référence n'ayant pas d'homologue dans la BDTOPO® (voir la Figure 36) illustre bien la complémentarité de ces sources.

Par exemple, Camptocamp a plusieurs catégories spécifiques aux sports de montagne comme l'escalade, les cascades ; ce sont des objets de repère moins représentés dans la base de données d'autorité BDTOPO®. En ce qui concerne la redondance des objets de repère entre les différentes sources de données, seuls 37 objets de repère homologues sont présents dans toutes les sources de données. Nous avons remarqué que les sources Camptocamp et Refuges.info sont les plus comparables avec la BDTOPO® en termes de représentativité des données : 1790 objets de repère de Camptocamp (59%) et 352 objets de Refuges.info (73%) ont un objet de repère homologue dans la BDTOPO®. Ce n'est pas le cas pour la source Geotrack où seulement 127 objets (11%) ont un objet homologue dans la BDTOPO®. Ceci montre la complémentarité des sources, qui est une richesse dans notre cas.

Pour analyser la redondance et la complémentarité des sources concernant les itinéraires, la mesure de connectivité d'un graphe d'itinéraire a été utilisée. En examinant séparément les différents réseaux des itinéraires pour les trois sources de données, on obtient 58 routes connectées pour Visorando, 28 pour Camptocamp et 28 pour Geotrack. Cette répartition signifie que les sources de données couvrent une grande superficie. En calculant ce même indicateur, mais avec les graphes fusionnés deux par deux, le nombre de composants connectés reste stable, à savoir 59 pour la fusion

entre Visorando et Camptocamp et 58 pour celle entre Visorando et Geotrack. Avec la fusion des sources, le nombre de tronçons augmente par composant connecté, mais la couverture ne varie pas. Les itinéraires des différentes sources ne se sont pas toujours connectés avec la fusion, ce qui nous mène à conclure que les différentes sources d'information sont non-redondantes.

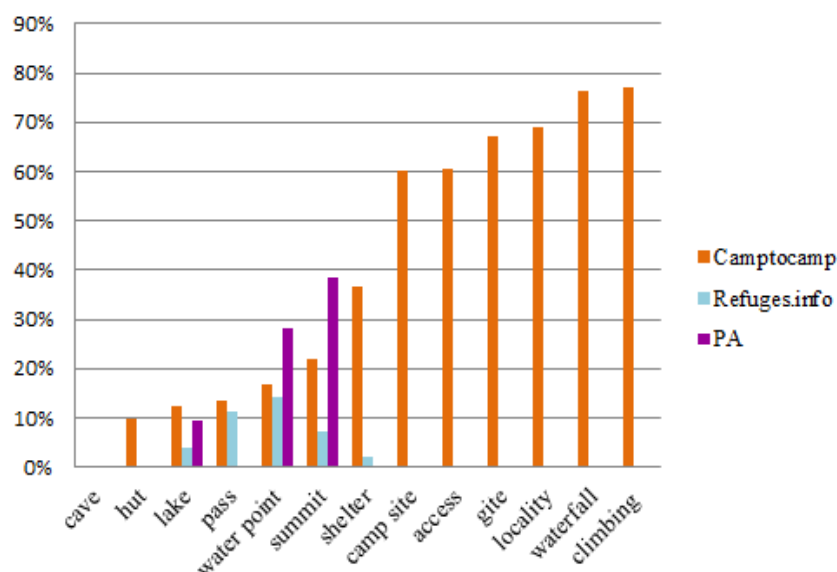


Figure 36. Distribution des points de repère non appariés par catégorie ; issue de (Van Damme et al., 2019)

Malgré cette complémentarité riche, nous avons constaté également une redondance. L'analyse de la redondance (liens 1 :1 validés) nous permet d'étudier l'hétérogénéité sémantique des différentes sources et de dériver les règles de correspondance sémantique utiles pour la constitution et l'interrogation d'un référentiel métier d'objets de repères et d'itinéraires. Si la victime exprime « Je suis à gauche du rocher du Vent », il est primordial de savoir que dans Camptocamp il y a une confusion entre les rochers et les sommets et que l'interrogation doit se faire non seulement auprès des objets de la catégorie « rocher », mais aussi de ceux de la catégorie « sommet ».

Ainsi, nous avons défini des matrices de confusion à la fois pour les itinéraires (niveau de difficulté) et pour les objets de repère (nature des objets).

Pour les itinéraires, la matrice de confusion des niveaux de difficultés entre les itinéraires Visorando et les itinéraires de Géotrack a été calculée. Chaque source contient cinq niveaux, les noms des niveaux sont différents, mais les paramètres considérés sont similaires (*e.g.* effort physique, mental et différence d'altitude). Dans l'ensemble, les deux sources sont en désaccord. Par exemple, le niveau moyen est réparti sur les niveaux facile, moyen et difficile dans les deux cas. Plus précisément, le nombre d'itinéraires de niveau facile dans Visorando prend toutes les valeurs des niveaux des itinéraires des zones protégées. Un parcours de niveau difficile est classé dans le niveau facile et vice-versa. Ce travail préliminaire devrait être examiné plus en détail à travers des stages encadrés par ma collègue Marie-Dominique Van Damme.

Concernant les objets de repère, nous avons pu remarquer que les sommets dans Camptocamp sont souvent appariés avec des sommets (547 objets), des pics (259 objets), des crêtes (29 objets), des rochers (29 objets). Une autre règle pertinente concerne les objets de repère classés dans la catégorie « accès ». Ces objets de repère sont appariés à des objets de type « stationnement » (40) ou aux cols (33). Deuxièmement, pour Refuges.info, on peut remarquer que les passages sont toujours appariés à des cols. D'autres appariements intéressants sommets/pics (16 objets de repères) et les objets d'eau/chutes d'eau (5) ont été trouvés. Un objet de type « point d'eau » dans Refuge.info représente une source d'eau mais aussi des objets tels que des ponts. Ces objets de repère représentant des ponts n'ont pas été appariés. Enfin, les résultats de l'analyse de la matrice de confusion Geotrack-BDTopo sont très surprenants. Par exemple, les objets de type « géologie »,

« histoire » sont associés à des canyons, des cols, des grottes et des abris. Cela montre la complémentarité des données de la source Geotrack par rapport aux autres sources. De plus, les points de vue sont souvent associés à des sommets et des cols.

La méthode d'appariement que nous avons utilisée apparie une source de référence à une source de comparaison. La cardinalité des liens appariés correspond à 1:1 ou 1:n (par post-traitement). Pour n sources d'information et sans source de référence connue *a priori*, nous devons exécuter l'algorithme $n*(n-1)$ fois, ce qui n'est pas une solution optimale étant donné notre application. Ainsi, une première perspective est d'adapter la méthode existante pour proposer une méthode d'appariement multi-sources. L'appariement des itinéraires nous a permis d'identifier le besoin d'améliorer la définition de ce que l'on appelle un itinéraire d'une part, et deux itinéraires homologues d'autre part. Nous pourrions par exemple considérer que deux itinéraires sont homologues s'il y a plusieurs objets de repère identiques et si les caractéristiques du terrain sont similaires entre deux objets de repère. Ceci nécessite bien évidemment de définir des liens entre les objets de repère et les itinéraires ainsi que d'identifier les types de relation de localisation telles que « voir » (e.g. le sommet est vu depuis l'itinéraire), « longer » (e.g. l'itinéraire longe le lac).

D'une part, ce travail nous a permis de constater que les sources que nous avons identifiées sur le web sont des sources très riches et nous avons prouvé leur potentiel pour la constitution d'un référentiel métier pour la localisation des victimes en montagne. De nombreuses questions de recherche se posent pour l'intégration de ces sources. Quel modèle pour le référentiel métier, étant donné que l'actualité et la qualité des données est primordiale ? Comment gérer la redondance et l'hétérogénéité des sources afin de ne pas perdre la richesse de chaque source ? Quel type de représentation favoriser pour les objets homologues ? Plusieurs solutions telles que la représentation multiple ou la représentation floue peuvent être envisageables. De plus, les objets de repère ne sont pas que des objets ponctuels, mais peuvent être aussi linéaires ou surfaciques. Le choix de la représentation n'est pas anodin, car la détermination d'une zone de localisation compatible en dépend (Bunel et al., 2019).

D'autre part, comme évoqué ci-dessus, grâce à l'appariement, nous avons pu identifier de nouveaux types d'objet, complémentaires de ceux existant dans les sources d'autorité, qui ont permis d'enrichir l'ontologie OOR (Olteanu-Raimond et al., 2020 b), ainsi que des liens de correspondance entre les types d'objet provenant de différentes sources. Les liens de correspondance doivent être formalisés et intégrés dans les ontologies OOR et OOI qui guideront l'interrogation du référentiel des objets de repère et itinéraires.

Chapitre 5. Conclusion et perspectives de recherche

Dans ce mémoire, j'ai présenté une synthèse des travaux de recherche autour de l'information géographique volontaire, en allant des mécanismes collaboratifs qui permettent la collecte jusqu'à son utilisation pour des besoins opérationnels en passant par l'évaluation et l'amélioration de la qualité. Ce chapitre synthétise les contributions significatives et ouvre des perspectives de recherche à court et long termes.

5.1 Information géographique volontaire : des pratiques courantes aux recommandations

Depuis une quinzaine d'années, le collaboratif a pris une place importante dans différents domaines pour répondre à la fois aux besoins sociétaux, commerciaux, environnementaux pour ne citer que ceux-ci. Quinze ans après les premières recherches autour de l'information géographique volontaire dans le domaine de la géomatique, nous pouvons conclure que le collaboratif n'était pas un effet de mode, au contraire il se développe de plus en plus.

Terminologie et initiatives. Le travail réalisé au sein de l'Action *COST Mapping and Citizen Sensors* a permis de générer des connaissances sur la terminologie et la chronologie des termes utilisés d'une part, et sur les différentes initiatives collaboratives, d'autre part. Deux contributions de type *génération de connaissances* méritent d'être citées :

- Proposition d'une classification de l'information géographique produite par la foule selon deux axes : un premier qui qualifie l'acte de la contribution (*i.e.* active/passive) et un deuxième qualifiant le type d'information collectée (*i.e.* géographique /non géographique).
- Une cartographie des initiatives collaboratives analysées selon sept critères : le thème, la nature et le type de l'IGV, l'expertise et la formation, l'accessibilité aux données et aux métadonnées, la qualité et l'utilisabilité, les informations concernant les contributeurs, les incitations à la participation. Cette analyse nous a permis d'identifier les bonnes et les moins bonnes pratiques et de définir des recommandations.

Usage de l'IGV au sein des ANC. Le travail que j'ai coordonné dans le cadre du groupe de travail WP3 de l'Action *COST Mapping and Citizen Sensors* a permis de générer des connaissances sur l'usage de l'IGV dans les ANC. Les contributions sont de l'ordre de la génération des connaissances :

- Cartographie des initiatives collaboratives menées par les ANC. Nous avons constaté que cette pratique était récente, que très peu d'ANC d'Europe utilisaient l'IGV et que la plupart ont proposé leur propre plateforme de collecte.
- Identification des obstacles à l'usage de l'IGV au sein des ANC. Nous avons distingué cinq obstacles : la qualité et la validation des données, les questions juridiques de droit des données et de responsabilité civile, l'expertise et la motivation de la foule, la pérennité des initiatives collaboratives autour d'une communauté stable de contributeurs, les craintes en matière d'emploi.
- Identification des éléments permettant d'aller au-delà de ces obstacles. D'une part, nous avons identifié les éléments qui permettraient de bâtir une relation de confiance et qui soit « gagnant-gagnant » entre la communauté des ANC et la communauté des contributeurs. D'autre part, nous avons défini des questions de recherche dont la communauté des chercheurs pourrait s'emparer pour réduire les obstacles identifiés.

Protocole de données et recommandations. L'existence d'un protocole de données ou de recommandations nous a semblé une nécessité pour toute initiative de collecte d'IGV. Cependant, il est très difficile de trouver un compromis entre un protocole strict qui briserait la motivation et la participation, et un protocole laxiste qui permettrait l'interprétation et la subjectivité, mais qui engendrerait une hétérogénéité au sein de l'IGV collectée. Les contributions marquantes sont :

- Proposition d'un protocole de collecte d'IGV vectorielle qui se veut générique et qui peut être implémenté selon les besoins et les objectifs du projet collaboratif à mettre en place.
- Liste de recommandations pour développer davantage le collaboratif au sein des ANC. Parmi celles-ci nous pouvons rappeler : développer ses propres plateformes collaboratives autour d'une communauté impliquant les organismes publics, les acteurs de l'éducation et les citoyens, créer une couche citoyenne, ajouter une plus-value en qualifiant l'information collectée et en la rendant compatible avec les référentiels d'autorité.

Plateforme collaborative : PAYSAGE. Dans le cadre du projet LandSense, nous avons instancié le protocole générique dans un contexte institutionnel pour la mise à jour de l'occupation du sol. Les contributions marquantes sont de l'ordre conceptuel et opérationnel :

- Proposition du concept de campagne décliné sous trois formes : opportuniste, guidée et flash, mise en œuvre à travers un outil web et une application mobile.
- Proposition de types de tâche collaborative : édition, classification, validation, ajout de nouveaux attributs, visualisation des activités des campagnes en cours.
- Proposition de types de contribution : web et *in-situ*.
- Identification des rôles des utilisateurs : l'administrateur, l'expert et le contributeur.

Pistes de recherche et améliorations

Vers une veille continue, technique et méthodologique concernant les pratiques collaboratives

La recherche réalisée sur la terminologie et les initiatives ne présente que l'état des pratiques et des initiatives dans la fenêtre temporelle 2005 – 2016. Elle nous a permis de mieux comprendre les concepts qui gravitent autour du collaboratif et de l'IGV et ainsi de dresser des recommandations qui visent la collecte, la gestion, la qualification et l'usage de l'IGV. Néanmoins, les pratiques du collaboratif et les initiatives évoluent, et la législation, avec notamment le règlement général sur la protection des données (RGPD)²² en vigueur depuis le 23 Mai 2018, restreint certaines pratiques telles que l'analyse des contributeurs, la collecte et l'utilisation des informations personnelles. Dans cette évolution constante, il est donc important de continuer à analyser l'évolution du collaboratif comme l'ont fait récemment (Yan et al., 2020), mais aussi de remettre en question certaines pratiques et recommandations.

Les initiatives collaboratives au sein des ANC ont également évolué. Le workshop Crowdsourcing in National Mapping organisé par EuroSDR du 16 au 17 Janvier 2020²³ a montré l'avancement des ANC par rapport aux initiatives collaboratives. La totalité des ANC représentées au workshop, excepté une agence, ont un programme collaboratif. Pour certaines, le collaboratif évolue à travers des cas d'étude, pour d'autres le processus est rodé pour certains thèmes, mais pour la totalité, le passage à l'échelle reste encore un verrou à lever. Parmi les obstacles que nous avons identifiés dans nos travaux (Olteanu-Raimond et al., 2017a), nous avons pu remarquer que le collaboratif n'est plus considéré comme une source de perte d'emploi. En revanche, la qualification de l'IGV et la motivation des contributeurs restent encore un défi pour les ANC. Nous avons remarqué également que très peu d'ANC impliquent les citoyens dans leur initiatives, préférant plutôt une collaboration avec des autorités locales. Les recommandations d'impliquer les associations sportives à travers une couche citoyenne (comme par exemple dans le prototype Map Gretel de l'ANC Finlandais (Rönneberg et al., 2019)) ou le système éducatif (voir les expériences des ANC de Finlande et des Pays-Bas ainsi que notre expérience dans PAYSAGES) restent de bonnes pratiques, à mon sens. Bien que nous puissions nous interroger sur le caractère « volontaire » de l'usage de ces applications, car les contributeurs (étudiants, élèves) font partie d'un programme sélectionné par leurs enseignants, cela permet au plus jeune public de se familiariser avec l'IG, et d'être sensibilisé à la qualité et aux problématiques sociales et environnementales. Cependant, l'objectif principal doit rester celui de l'apprentissage à travers la pratique et leur implication doit être pédagogique. Ce travail commun entre les ANC et la communauté de recherche est très bénéfique pour les deux communautés à mon sens, car il permet de partager les avancées au sein des deux communautés et de lever les verrous qui existent encore. C'est pour cela que dans le cadre d'une collaboration LandSense – EuroSDR, j'ai initié avec Joep Cromptvoets l'organisation d'un workshop qui aura lieu fin 2020 sur l'usage de l'IGV

²² <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

²³ <http://www.eurosd.net/workshops/eurosd-workshop-crowdsourcing-national-mapping>

pour la production et la gestion des données d'occupation du sol et qui impliquera la communauté des chercheurs en IVG et en traitement d'image et les ANC.

La mise en œuvre des propositions à travers la plateforme PAYSAGES et les campagnes réalisées nous ont permis de tirer quelques leçons et d'identifier les verrous qui restent à lever.

Vers une réflexion pour la connexion à une plateforme collaborative.

Une recommandation à remettre en cause est la connexion. En effet, nous avons recommandé plusieurs types d'identification allant d'une simple identification (email, mot de passe), jusqu'à une identification sécuritaire en utilisant le numéro de CNI pour certains types de thématique). Certaines recommandations ne sont plus d'actualité, étant donné le nouveau règlement RGPD. D'autres, bien que respectant la réglementation en termes de protection de la vie privée, se sont avérées peu adaptées. Par exemple, le retour d'expérience de l'ANC de Finlande montre que l'identification avec le numéro de CNI a été adaptée pour les partenaires publics mais a été un frein à la participation par les autres catégories de contributeurs²⁴. Dans la plateforme PAYSAGES, nous avons eu des retours de contributeurs mentionnant l'inconvénient de créer un nouveau compte pour pouvoir participer. L'expérience nous mènera peut-être à revoir ces recommandations et à nous questionner sur la connexion. Une solution envisageable serait le système d'authentification fédérée qui implique que les utilisateurs se connectent avec des comptes déjà existants. Bien que complexe et difficile à mettre en place, cela semble convenir aux observatoires citoyens (Matheus et Batič, 2016). Son application dans un contexte institutionnel reste encore à étudier.

Vers une meilleure connaissance des communautés collaboratives et le rôle qu'elles peuvent avoir dans une plateforme collaborative

Une des recommandations que nous avons implémentées dans la plateforme PAYSAGES a été le recueil optionnel des informations permettant de caractériser le profil des contributeurs telles que la profession, la catégorie d'âge, etc. Cette recommandation est donc à revoir également. En effet, dans le cadre de la plateforme PAYSAGES, au moment de la création d'un compte, les contributeurs avaient la possibilité de remplir des informations supplémentaires concernant leur profil. Ces informations supplémentaires, bien que respectant le règlement GDPR, n'ont quasiment pas été remplies par nos contributeurs. Cependant, étant donné que nous avons organisé des campagnes flash impliquant un profil de contributeurs à la fois, nous avons été en mesure d'analyser le processus de collecte et l'IGV collectée selon l'expertise du contributeur.

Dans sa thèse, Flora Hayat (Hayat, 2019) fait une analyse intéressante de l'éco-système de la plateforme OSM et décrit trois directions possibles au regard de la communauté OSM : (1) la poursuite de l'amélioration de la qualité de la base de données, (2) OSM source de renouvellement et (3) OSM source unificatrice. Si ces directions sont tout à fait plausibles dans certains contextes, je m'interroge également sur une quatrième direction, à savoir OSM agrégateur de données, et sur l'impact que cela pourrait avoir sur les contributeurs de la communauté OSM. Ces imports massifs peuvent avoir un effet sur la démotivation des contributeurs, comme cela a été le cas lors de l'import de la base de données Tigger dans OSM. En partant du constat qu'il y a de plus en plus d'imports massifs qui sont réalisés, une première perspective de travail serait d'établir un état de l'art sur les sources d'information importées massivement dans OSM et d'analyser l'évolution de ces données au cours du temps et le comportement des contributeurs face aux imports.

24

http://www.eurosdrr.net/sites/default/files/images/inline/crowdsourcing_projects_in_the_national_land_survey_of_finland_risto_ilves_nls.pdf

Vers une meilleure stratégie d'animation

L'animation d'une plateforme ne doit pas être sous-estimée et reste un défi à lever. En effet, dans le monde idéal, tout organisme ou association proposant une plateforme souhaiterait qu'il y ait beaucoup de contributeurs, qu'il y ait une participation de longue durée, en quantité et de qualité. Cependant, cela reste un rêve difficile à atteindre dans la réalité. L'animation à long terme est un défi à lever pour tout projet collaboratif. Notre première expérience nous a montré que le concept de campagne pourrait être un élément bénéfique pour l'animation d'un projet collaboratif. Notamment, les campagnes flash et guidée nous semblent adaptées à long terme.

Nous avons proposé deux types de contribution : web et *in-situ*. Nous avons remarqué une différence entre les deux types de contribution à la fois en termes d'animation et de motivation. La campagne flash web est plus conviviale et permet aux participants de créer des liens entre eux et d'opérer une aide permanente auprès des participants, en cas de besoin. Les campagnes flash *in-situ* nécessitent une meilleure préparation en amont de la zone visitée. Lorsque les contributeurs en ont eu le choix, à la fois les étudiants et les experts ont préféré travailler en binôme. En termes d'efficacité, il n'y a pas de doute, plus de contributions sont collectées pendant la campagne flash web. Cependant, tout ceci est à nuancer, car l'information terrain est parfois indispensable. Une piste d'amélioration serait de mieux coupler les campagnes web et guidée, c'est-à-dire de réaliser d'abord une campagne web, durant laquelle des cas difficiles pourraient être remontés, qui seraient ensuite visités sur le terrain.

Enfin, nous avons appris lors de la première campagne de 2018 qu'il faut éviter de définir un objectif général avec des tâches parallèles. Le contributeur peut être dérouté et cela freine la première participation. Il s'est avéré être une bonne pratique de se concentrer sur un objectif spécifique et de le diviser en différentes tâches qui peuvent être abordées à des périodes différentes et avec différents profils de contributeur. Nous avons mis en place cette stratégie lors de la deuxième vague de campagnes réalisées en 2019 en ne donnant accès aux contributeurs qu'à une campagne à la fois.

Vers une meilleure plateforme centrée-utilisateur

- Impliquer les contributeurs dès la phase de conception d'une initiative collaborative.

En effet, les travaux de recherche sur la motivation montrent que les contributeurs participent s'ils y trouvent un intérêt. Dans un projet collaboratif proposé par une ANC, la motivation est plus difficile à identifier. Dans les travaux réalisés, un des profils de contributeurs sont les partenaires de l'IGN, utilisateurs des référentiels produits par l'IGN. Pour nous assurer de leur participation, nous les avons interrogés sur leurs besoins et les avons impliqués dans la phase de test des outils. Cependant, il est parfois difficile de trouver un compromis et d'adapter les fonctionnalités en fonction des besoins de chacun.

- Adapter les tâches en fonction de l'expertise des contributeurs.

Dans le cas de l'occupation du sol, certains profils de contributeur ont trouvé la nomenclature complexe, d'autres en revanche auraient souhaité une nomenclature plus détaillée. Ainsi, une perspective de recherche serait d'identifier les tâches selon les différents profils d'utilisateur et leurs motivations (Budhathoki et Haythornthwaite, 2013; Bucher et al., 2016). Combiner le concept de campagne flash/guidée en se focalisant sur un besoin spécifique et par type de profil nous semble une piste qui mérite d'être davantage explorée. Cela nécessitera donc plus de genericité au niveau des fonctionnalités, d'identifier plus de tâches et les rôles de chaque profil de contributeur.

- Développer davantage les rôles d'administrateur et d'expert.

Dans la plateforme PAYSAGES, les fonctionnalités de ces rôles n'ont pas été implémentées par faute de temps ; les rôles d'administrateur et d'expert ont été tenus par certains membres du projet et la mise en place de différentes campagnes a été réalisée manuellement (import/export de données,

fonctionnalités activées/inactivées, choix des nomenclatures, etc). Or, nous pensons que ces deux rôles permettraient une meilleure adaptation aux besoins et aux profils des contributeurs.

Pour avancer sur ces aspects d'interface, à savoir l'adaptation aux besoins des utilisateurs et l'évaluation des tâches, il est nécessaire de se rapprocher de la communauté de recherche en géovisualisation, notamment les chercheurs de l'axe Géovisualisation du LASTIG (Christophe, 2017; Touya et al., 2017) ou de l'Action Prospective de GDR MAGIS Cartographie & Cognition²⁵.

5.2 Qualification et amélioration de l'information géographique volontaire

Le premier constat concernant la qualification et l'amélioration de l'IGV est le caractère *ad-hoc* des méthodes proposées. En effet, elles dépendent du type d'IGV et de la manière dont elle a été collectée afin de mieux prendre en compte les hétérogénéités et le biais que la collecte de données peut engendrer.

Qualification des alertes opportunistes

Nous avons proposé une méthode permettant de déterminer la confiance des alertes opportunistes (*i.e.* remontées par différents partenaires de l'IGN via la plateforme Espace collaboratif (ancien RIPART)) afin de faciliter le contrôle qualité. Il s'agit d'un type d'IGV peu rencontré dans le domaine du collaboratif en dehors des initiatives proposées par les ANC. Notre contribution, d'ordre méthodologique, est la suivante :

- Associer aux commentaires et aux thèmes des alertes un ensemble de concepts provenant d'une taxonomie de concepts géographiques.
- Proposer une mesure de confiance basée sur un ensemble d'indicateurs analysant la cohérence des alertes par rapport à leurs propres caractéristiques ou à leur contexte spatial.
- Prise en compte des approches « géographique » et « intrinsèque » pour dériver une confiance qualitative dans les alertes.
- Décision multi-niveaux : à la fois la confiance qualitative finale et la confiance exprimée à travers les indicateurs de cohérence peuvent être utilisées.

Qualification des alertes guidées

Pour répondre aux objectifs d'amélioration et de mise à jour des données d'occupation de sol, nous avons généré deux types d'alerte : les alertes de changements venant d'un service de détection de changement proposé dans LandSense et les alertes correspondant à un ensemble d'objets géographiques ayant des classes d'usage agrégées, ou qui sont susceptibles d'avoir changé. Cette dernière catégorie d'alertes a été générée par nos soins. Nos contributions d'ordre conceptuel et méthodologique et sur la génération de connaissances sont les suivantes :

- Proposition d'une stratégie collaborative pour la qualification des deux types d'alerte : campagnes flash, web et in-situ adaptées aux profils des utilisateurs et aux tâches à réaliser ;
- Proposition d'une méthode pour valider les changements générés automatiquement et pour créer un jeu de données de référence en suivant les recommandations de (Olofsson et al., 2014) ;
- Adaptation de l'indicateur Kappa de Fleiss (1971) pour mesurer l'accord des contributeurs lorsque l'IGV multiple est disponible.
- Génération de connaissances concernant les comportements de différents profils de contributeur face à la tâche de validation et la création d'un jeu de données de référence.

²⁵ <http://gdr-magis.imag.fr/actions-prospectives/geo-visualisation-cognition/>

Qualification et amélioration des traces GNSS

Concernant la qualification et l'amélioration des traces GNSS, notre contribution est méthodologique et implique deux aspects : l'amélioration et la qualification.

- Améliorer la géométrie des traces pour mieux s'approcher de la représentation d'un objet géographique linéaire représentant une route et s'abstraire du comportement de l'individu. Ainsi, nous avons proposé deux méthodes permettant le filtrage de ce que nous considérons, dans notre contexte applicatif, comme du bruit.
 - La méthode de détection de comportement secondaire se caractérise par sa simplicité et est basée uniquement sur la géométrie des traces en distinguant les traces qui suivent le même itinéraire à l'aller et au retour.
 - La méthode de détection de points GNSS aberrants est basée sur l'apprentissage supervisé et permet de détecter les points GNSS aberrants dans des traces hétérogènes caractérisées par : le manque de métadonnées, l'incomplétude des valeurs d'attribut, le peu d'information sémantique et le peu de redondance. De plus, étant donné que les règles apprises sont basées uniquement sur les indicateurs intrinsèques, elle peut être utilisée en absence d'IG d'autorité. Une contribution sous-jacente est l'ensemble des indicateurs intrinsèques et extrinsèques proposés pour prendre en compte les facteurs qui peuvent influencer le processus d'acquisition GNSS.
- Qualifier l'exactitude de points GNSS composant une trace. La méthode proposée s'appuie sur la même méthode que la détection des points GNSS aberrants et les mêmes indicateurs intrinsèques et extrinsèques. Par contre, différentes stratégies de définition de données d'apprentissage ont été proposées. L'exactitude des points GNSS (qualitative et binaire) est évaluée grâce aux règles apprises. La méthode proposée s'affranchit de l'usage de l'IG d'autorité homologue mais a besoin de l'IG contextuelle.

Pistes de recherche et améliorations

Vers une meilleure qualification des alertes opportunistes à travers...

- **l'amélioration de la définition du contexte spatial**

L'identification du contexte spatial des alertes présente deux limites. La première est le fait que le contexte spatial est basé seulement sur des relations spatiales d'inclusion et de proximité en utilisant des zones tampon. Afin de remédier à cette limite, nous pourrions utiliser d'autres relations topologiques ou des relations spatiales (*e.g.* à gauche, en face, à côté) provenant des commentaires (*e.g.* la route à gauche de la maison de quartier a été modifiée). La deuxième limite est due au fait que seul le référentiel topographique de la BDTOPO® a été utilisé pour déterminer le contexte spatial de l'alerte. Par conséquent, certaines alertes n'ont pas de contexte spatial identifié car il n'y a pas d'objet géographique à proximité. Pour remédier à cette limite nous pourrions, par exemple, nous appuyer sur les données d'occupation du sol. Pour éviter de générer des incohérences entre les données topographiques (géométrie discontinue) et les données d'occupation du sol (géométrie continue), une piste de recherche serait de définir deux contextes spatiaux indépendants pour chaque alerte, puis de les combiner.

- **une meilleure combinaison des indicateurs**

La combinaison des indicateurs que nous avons proposée consiste à définir des règles basées sur des seuils choisis empiriquement et à dériver une confiance mesurée qualitativement. Avoir une mesure qualitative facilite, en effet, la prise de décision. Par contre, dans notre proposition la combinaison des indicateurs n'est pas assez précise. En effet, nous avons un qualificatif pour chaque indicateur (oui, non, peut-être) et la combinaison des indicateurs pour avoir la confiance totale est basée sur la

majorité. Pour remédier à cette limite et avoir une modélisation plus fine, une piste de recherche serait d'envisager de définir un système d'inférence floue comme proposé par exemple dans les travaux de (Yan et al., 2016). Ainsi, pour chaque indicateur, il serait possible d'inférer un niveau de cohérence en utilisant les degrés d'appartenance. De cette manière, l'étape de combinaison des différents indicateurs est plus précise et reste quantitative. Le passage d'une valeur quantitative à une valeur qualitative est envisageable en définissant des règles de défuzzification.

- une meilleure automatisation pour une meilleure ré-utilisation

Une autre limite de l'approche est le manque d'automatisation. En effet, plusieurs indicateurs nécessitent un travail préparatoire manuel ou semi-automatique. Par exemple, la taxonomie mise à notre disposition n'a pas été exploitée à sa juste valeur. A la fois pour les commentaires et les thèmes, la liste des concepts associés aux mots significatifs et aux thèmes a été réalisée manuellement, en amont. Une piste de recherche est de pouvoir la générer à la volée. Cela nécessite d'évoluer ou adapter la taxonomie existante en modélisant les nouvelles connaissances acquises et en définissant d'autres relations que « isA » entre les concepts de la taxonomie.

Les commentaires se sont avérés être une source riche d'information. Cependant, le traitement des commentaires a été également réalisé manuellement. Des techniques de traitement du langage qui nous permettraient de déterminer des mots signifiants pourraient être utilisées en employant un vocabulaire contrôlé. La difficulté est de définir ce vocabulaire contrôlé dans un contexte où le nombre de commentaires est faible. Une piste de recherche serait de s'inspirer de la liste des erreurs que nous pourrions trouver dans le wiki OSM.

Vers une amélioration de l'évaluation de l'exactitude des points GNSS

Même si la performance de l'approche est modérée, nous avons démontré son potentiel pour générer des alertes de mise à jour pour les sentiers. De plus, elle ouvre de nouvelles perspectives pour l'évaluation de la qualité des traces GPS, indépendamment du réseau routier référentiel.

Plusieurs pistes d'amélioration pourraient être envisagées. Premièrement, nous pensons que les résultats pourraient être améliorés en ajoutant des indicateurs tels que ceux basés sur la constellation des satellites ou la connaissance du dispositif GPS. Pour cela, nous pourrions adapter la méthode sur un jeu de traces GNSS collectées en utilisant un protocole de collecte qui sauvegarde les métadonnées. Une deuxième limite est l'évaluation booléenne. Elle pourrait être remédiée en s'appuyant sur une approche basée sur les règles floues (Yan et al., 2016). Ainsi, de la même manière que pour la qualification des alertes opportunistes, nous pourrions définir plusieurs niveaux d'exactitude (très forte, forte, moyenne, faible, très faible) et générer un degré d'appartenance à ces classes de valeurs en utilisant les règles et les seuils appris grâce à l'approche proposée par S. Ivanovic. Ces différentes valeurs issues de chaque indicateur pourraient être combinées grâce aux règles conjonctives afin de dériver une exactitude finale.

Vers une généricité et transférabilité des méthodes de qualification des traces GNSS

Les méthodes d'identification du comportement humain secondaire et des points GNSS aberrants ont été appliquées aux traces GNSS collectées en zone de montagne.

Etant donné que la méthode de filtrage des comportements humains secondaires n'est basée que sur la géométrie, nous estimons qu'elle devrait être applicable à divers types de comportement humain secondaire, indépendamment des caractéristiques de l'environnement. Les comportements humains secondaires peuvent être considérés comme des stations (*i.e.* des lieux où les individus s'arrêtent pour effectuer une activité). Détecter et analyser les stations dans les trajectoires individuelles est un objet de recherche lorsqu'on s'intéresse aux déplacements individuels (Buard, 2011; Olteanu-Raimond et al., 2011). Ainsi, notre méthode pourrait être utilisée pour détecter des stations dans les trajectoires spatio-temporelles hétérogènes. Des recherches supplémentaires seraient nécessaires pour déterminer les limites de la station et d'autres informations sémantiques telles que le type d'activité ou la durée de stationnement. Cette dernière est difficile à estimer pour

les traces dont il manque l'estampille temporelle. L'utilisation du contexte spatial ou la mise en correspondance des traces avec leur description textuelle, le cas échéant, peuvent également être explorées, comme dans les travaux de Medad et al. (2018).

Sans l'accès aux métadonnées, il est difficile de savoir pour chaque trace le type de capteur utilisé (e.g. GPS pro ou intégré dans un smartphone). Il en va de même pour le type d'activité sportive (e.g. course à pied, vélo, randonnée). Toutefois, l'analyse des sites web et des traces elles-mêmes nous a permis d'observer que différentes activités étaient pratiquées dans notre jeu de traces GNSS. Par conséquent, nous considérons que ces deux méthodes sont indépendantes du type de capteur et de l'activité exercée ainsi que du mode de transport (e.g. la marche ou le vélo).

Enfin, en ce qui concerne la méthode de détection des points aberrants, d'un point de vue technique, la méthode est applicable aux traces GNSS collectées dans d'autres environnements tels que les zones urbaines ou rurales. Toutefois, il est prévisible que ses performances varient dans un environnement opposé tel que la zone urbaine. Ceci est dû au fait que les règles sont apprises en utilisant des traces GNSS en zone de montagne qui sont différentes de celles en ville. Par exemple, en zone urbaine il y a plus d'obstacles, donc l'effet de canyon urbain devrait être plus important, puisque les déplacements en ville sont plus contraints par le réseau routier, peu d'itinéraires « hors route » étant possible. Afin d'évaluer la transférabilité de la méthode, une piste de recherche est d'appliquer la méthode sur des jeux de traces GNSS situés dans différents contextes (e.g. urbain, rural, hautes montagnes, etc.). Nous pourrions nous inspirer de la méthode d'évaluation suivie par Truong (2020) pour détecter le vandalisme dans les données OSM dans différents environnements spatiaux.

Vers une qualification globale à l'échelle de la trace

L'approche d'évaluation des traces GNSS est locale, à savoir à l'échelle du point GNSS, et a été transférée à l'échelle du segment. Pour aller plus loin, une piste de recherche envisageable serait de proposer un ensemble d'indicateurs à l'échelle de la trace basés sur des caractéristiques intrinsèques telles que la résolution géométrique, l'exactitude géométrique, l'hétérogénéité géométrique, l'hétérogénéité des vitesses lorsque l'estampille temporelle est disponible, etc. Ce travail pourrait s'appuyer sur des méthodes statistiques permettant de mesurer l'homogénéité, l'hétérogénéité, la dispersion intra-trace ou les profils moyens ou médians, si plus de redondance (Etienne, 2011). La qualification des traces à travers des indicateurs permettrait ensuite de filtrer les traces qui ne correspondent pas aux critères de qualité exigés par une application donnée.

Vers une meilleure qualification des alertes guidées

Le nombre de contributeurs validant une alerte est un défi pour la qualification de l'IGV multiple. Il y a d'une part la variabilité, et d'autre part le nombre optimal de contributeurs qui se prononcent sur une alerte. Dans notre cas, et dans la collecte de l'IGV en général, il est difficile de maîtriser cet aspect, certaines alertes seront concernées par plusieurs observations (*i.e.* classification d'occupation du sol) tandis que d'autres feront l'objet d'une seule observation. Une des limites de notre approche est que les alertes ayant une seule observation n'ont pas fait l'objet de qualification. Le manque de qualification a été seulement modélisé dans le processus de fusion d'informations pour la mise à jour de l'OCS-GE. Pour remédier à cette limite, une piste de recherche est d'appliquer l'approche géographique proposée par Goodchild et Li (2012). Ce travail se rapprocherait du travail que nous avons réalisé pour la qualification des alertes opportunistes et nécessiterait de définir le contexte spatial adapté pour qualifier les alertes d'occupation du sol qui ne sont pas ponctuelles mais caractérisent des zones géographiques de taille différente. Une deuxième piste pourrait être d'organiser des campagnes *flash* uniquement pour les alertes n'ayant qu'une ou deux observations.

Pour déterminer le jeu de données de référence servant de classification de référence à partir d'observations multiples, nous avons appliqué une approche basée sur le vote majoritaire. Cette approche suppose que ce dernier est correct, que chaque alerte a le même niveau de difficulté et que chaque contributeur a la même expertise. Or, dans le cas d'IGV, ces hypothèses ne sont pas

toujours vraies. Ainsi, l'exactitude thématique calculée à partir du vote majoritaire peut être problématique. Une piste d'amélioration serait d'estimer l'exactitude à travers des mesures d'entropie qui sont facilement interprétables et peuvent également fournir des informations sur les performances des classificateurs. Ces mesures d'entropie sont utilisées avec succès au sein de la communauté des écologistes pour analyser le suivi de la biodiversité (Sherwin et Fornells, 2019).

5.3 Intégrer différentes sources d'information géographique volontaire et institutionnelle en prenant en compte leur hétérogénéité

Mise à jour du référentiel routier d'autorité

Nous avons proposé une méthode permettant d'identifier des mises à jour pour les sentiers d'un référentiel routier. Les contributions sont :

- Proposition d'une méthode d'appariement de données géographiques permettant de s'affranchir des caractéristiques thématiques, sémantiques ou topologiques nécessaires à aux approches d'appariement existantes dans la littérature. De plus la méthode intègre l'exactitude des points GNSS composant la trace GNSS.
- Proposition d'une méthode permettant de combiner différents critères caractérisant la qualité de la trace, son tracé par rapport à des sentiers du référentiel d'autorité et la redondance des traces.

Mise à jour de données d'occupation du sol

Nous avons proposé une méthode permettant de mettre à jour l'usage des classes des données d'occupation du sol. Les principales contributions sont :

- Proposition d'une méthode de fusion d'informations provenant de sources hétérogènes qui permet de distinguer un changement significatif et un changement non significatif d'une part, et une mise à jour et une mise à niveau, d'autre part.
- La prise en compte des spécificités de chaque source d'IGV, de la qualité de l'IGV lorsqu'elle était disponible et des spécifications à travers une approche basée sur la modélisation des connaissances grâce au formalisme de la théorie de Dempster-Shafer.

Constitution d'un référentiel métier

Le travail de recherche décrit en Section 4.4 représente une première étape dans la constitution d'un référentiel métier pour l'aide à la localisation de victimes en montagne. Ce travail contribue à la communauté scientifique à travers :

- Une ontologie d'application décrivant les types et les caractéristiques des objets de repère dont une première version a déjà été publiée ;
- Des connaissances générées concernant les sources d'IGV ou d'autorité contenant des objets de repère ou des itinéraires. Les connaissances sont de plusieurs ordres : les caractéristiques de l'information géographique partagée, leur modèle, la complémentarité et la redondance des sources ;
- Un algorithme de reconstruction des routes développé par mes collègues de LaSTIG pour remédier aux limites de l'algorithme de map-matching ;
- Un *plug-in* de téléchargement de l'IGV en utilisant différents API des sites web identifiés comme adaptés à notre besoin.

Pistes de recherche et améliorations

Aller au-delà de l'identification des alertes de mise à jour dans les référentiels topographiques.

Plusieurs pistes pour avancer sur cet axe de recherche peuvent être identifiées :

- Analyser tous les types de cardinalité des liens d'appariement du processus d'appariement. Dans les travaux de S. Ivanovic, nous nous sommes intéressés uniquement aux liens 1 :0 (une entité réelle n'est pas représentée dans la base de données géographique d'autorité). Cependant, les liens n :m permettent de confirmer que les objets géographiques existent toujours dans la réalité. En plus de la confirmation, le lien peut être utilisé pour réaliser un transfert d'attributs et enrichir le référentiel topographique. La difficulté ici est de bien gérer la problématique de niveaux de détails différents entre les différentes sources d'IG utilisées. Les liens 0 :1 peuvent être interprétés comme des objets géographiques de la base de données d'autorité à supprimer car ils n'existent plus dans la réalité. Ce cas, plus complexe à traiter, nécessite davantage des mécanismes de raisonnement sur le comportement de l'individu, du contexte spatial ainsi qu'une meilleure qualité et quantité d'IGV.
- S'appuyer davantage sur les spécifications des bases de données géographiques pour proposer des mises à jour qui respectent la définition, la sélection et la cohérence des objets géographiques, telles que définies dans les spécifications (Sheeren, 2005).
- Combiner l'usage de l'IGV avec d'autres sources d'IG. Une des sources d'IG qui nous semble pertinente est l'IG issue des traitements d'image et qui représente le changement. Dans le cadre du post-doc de Lanfa Liu, nous avons exploré cette piste en utilisant des cartes raster représentant les bâtiments détectés à partir des orthophotos. Les premiers résultats ont montré que cette IG couplée avec les spécifications du thème bâti peut être une source fiable pour la mise à jour du thème bâti. Nous souhaiterons continuer à explorer cette piste avec mes collègues de LaSTIG experts du domaine du traitement d'image (Gressin et al., 2014; Le Bris et Chehata, 2011).

Vers une mise à jour de données d'occupation du sol complète. Plusieurs pistes pour avancer sur cet axe de recherche peuvent être identifiées :

- Etendre la méthode actuelle pour mettre à jour les classes de couverture du sol. L'extension concerne l'alignement des nomenclatures de l'OCS-GE et des différentes sources d'IGV caractérisant les changements de couverture. En fonction des types d'IG disponibles, des pré-traitements supplémentaires pourraient être nécessaires.
- Etendre la méthode actuelle pour permettre la mise à jour des limites de classes. Ceci nécessite bien évidemment d'avoir de l'IGV de type polygone, mais également de prendre en compte la cohérence de la couverture et l'usage du sol après la mise à jour thématique.

Vers une approche basée sur les ontologies

La modélisation des connaissances représentant les objets de référence (les objets de repère, les itinéraires, les objets cartographiques) et les relations de localisation est réalisée sous la forme de différentes ontologies. Ce travail en cours, exposé en Section 4.4, continuera de la manière suivante :

- Faire évoluer l'ontologie des objets de repère afin de modéliser les caractéristiques des types d'objets de repère et de décrire les liens de correspondance entre les différents schémas de sources d'IG que nous avons identifiés comme ayant un fort potentiel pour le secours en montagne.
- Finaliser les ontologies d'itinéraires et des objets cartographiques qui modélisent des connaissances des objets non matérialisés sur le terrain. Pour le concept d'itinéraire, nous nous appuyons sur la proposition de Teulade-Denantes et al. (2015) qui définit un itinéraire comme un ensemble de sentiers ou de chemins à suivre et est porté par une infrastructure

de réseau physique composée de routes et de sentiers. La construction de l'ontologie des itinéraires suit la même méthode que l'ontologie des objets de repère en s'appuyant sur trois sources : les informations provenant des experts, les enregistrements d'alertes et les sources d'IG volontaire ou d'autorité. Enfin, pour l'ontologie des objets cartographiques permettant de modéliser les objets présents sur la carte (*e.g.* toponyme, limite administrative), nous utiliserons les enregistrements audio et nous pourrions nous appuyer davantage sur les travaux de Dominguez et al. (2012).

- Ontologie des relations de localisation. Dans sa thèse, Mattia Bunel propose une ontologie des relations de localisation qui identifie et hiérarchise les types de relation de localisation utilisés pour se localiser en montagne en utilisant les enregistrements d'alertes et en s'appuyant sur l'ontologie des relations spatiales GUM-3-Space proposée par Bateman *et al.* (2010). L'ontologie des relations de localisation permet d'identifier les méthodes de calcul des zones compatibles en fonction des indices reçus. Dans l'approche proposée par Mattia Bunel, un type de relation de localisation a au moins une méthode localisation. Différents défis restent à relever concernant le lien entre le type de relation de localisation et le type d'objet de référence, les hypothèses de spatialisation pour chaque indice et les paramètres à définir.

5.4 Directions de recherche à long terme

1. Vers une mise à jour des référentiels d'autorité plus automatisée

Les premiers enseignements de nos recherches effectuées nous montrent le potentiel de l'IGV pour la mise à jour de référentiels et met en lumière le caractère *ad-hoc* de la démarche qui doit être différente d'un type de référentiel à un autre. Par conséquent, l'étape à franchir consisterait à proposer un processus de mise à jour complet qui devrait aller de l'identification des mises à jour et jusqu'à l'intégration des mises à jour identifiées pour produire un référentiel, en passant par l'analyse de la cohérence avec les données elles-mêmes (pour la nouvelle version produite), les versions antérieures et les spécifications.

De nombreuses questions de recherche restent encore à être traitées telles que : Comment gérer les hétérogénéités géométriques, sémantiques et temporelles entre les différentes sources d'IGV et l'IG d'autorité ? Comment prendre en compte les spécifications ? Comment évaluer la qualité des mises à jour ? Doit-on prévoir une étape supplémentaire de vérification terrain ou peut-elle être réalisée par stéréorestitution ? Comment créer l'information géographique de référence qui nous permettra d'évaluer une partie de la qualité ? Doit-on envisager une autre démarche qualité pour un référentiel d'autorité mis à jour à partir d'IGV ?

2. Vers une production de référentiel OCS combinant l'approche raster et vecteur

Les coûts de production des données d'occupation du sol selon les méthodes traditionnelles, aussi bien à l'échelle européenne qu'à l'échelle nationale, sont très élevés. Ceci a un impact sur l'actualité (*e.g.* six ans pour Corine Land Cover, trois ans pour l'OCS-GE), sur la couverture de ces données (*e.g.* pour OCS-GE seuls quelques départements sont couverts) et sur l'exactitude sémantique (*e.g.* des classes d'usage et des couvertures mixtes). Dans ce contexte, nous souhaitons explorer une autre piste de recherche qui consisterait à s'éloigner de la méthode traditionnelle et de proposer un processus de production et de mise à jour des données d'occupation du sol à l'échelle européenne en s'appuyant sur des approches d'intégration multi-sources (*e.g.* images satellitaires multi-dates, images aériennes, IG d'autorité, IGV). L'objectif applicatif est donc de disposer de données d'occupation du sol plus actuelles, plus fines, plus fiables et multi-temporelles. Cette piste de recherche serait à explorer, entres autres, à travers le projet **Top-Land-IQ** (Training Of the next PhD generation in cutting-edge LANd cover and land use Data management, Integration, Quality-assurance, and application) qui a été soumis en janvier 2020 au programme H2020 Marie

Sklodowska-Curie Innovative Training Networks et auquel IGN participe en tant que partenaire. Parmi les 15 sujets de thèse traitant des objectifs scientifiques différents allant de la définition d'une nomenclature des données d'occupation du sol unificatrice jusqu'aux modèles de prévision ou de changements d'occupation du sol, je suis impliquée dans le co-encadrement de trois thèses en collaboration avec l'Université de Coimbra de Portugal et IIASA. Les objectifs scientifiques des sujets de thèse initiés sont :

- Proposer une méthode permettant de produire des données d'occupation du sol vecteur qui combinerait les approches raster/vecteur pour la définition des limites géométriques et des classes d'usage et de couverture du sol, et qui faciliterait la mise à jour annuelle.
- Utiliser différentes sources d'IGV (e.g. photographies géolocalisées, données OSM) pour améliorer les classes d'usage et de couverture.
- Proposer des méthodes pour la constitution et la mise à jour de données servant de « *vérité terrain* » basées sur le collaboratif afin de permettre d'évaluer la qualité des données d'occupation du sol ainsi générées.
- Proposer des indicateurs de qualité, autres que l'exactitude géométrique et sémantique, traditionnellement utilisés malgré leurs limites, et des approches pour communiquer la qualité des données produites d'une manière compréhensible pour les utilisateurs.

3. Vers la constitution d'un référentiel de données géographiques uni-source

Dans le cadre de la thèse d'Ibrahim Maidaneh Abdi, nous nous intéressons à la constitution d'un référentiel de données géographiques à partir de l'IGV et plus précisément à partir d'OSM. Il s'agit ici d'un objectif applicatif qui vise à remédier à la problématique du manque de référentiels dans les pays d'Afrique. Pour atteindre cet objectif, il est nécessaire d'avancer sur deux défis : la qualité de l'IGV en l'absence d'IG d'autorité et la définition d'une méthode permettant de passer du modèle OSM avec une nomenclature riche et hétérogène à un modèle qui se veut homogène et plus sélectif.

Qualifier l'IGV en absence de l'IG d'autorité. L'hypothèse explorée dans la thèse d'Ibrahim Maidaneh Abdi est d'analyser la possibilité de dériver l'exactitude géométrique (position, forme) de l'IGV à partir des indicateurs intrinsèques permettant de caractériser les erreurs géométriques (Girres, 2012; Follin et al., 2019). Sa démarche combine les approches extrinsèque et intrinsèque et est basée sur l'apprentissage automatique, comme par exemple dans les travaux de Mohammadi et Malek (2015) et Xu et al. (2017). Plus précisément, elle consiste premièrement à calculer les **indicateurs extrinsèques** grâce à un appariement entre l'IGV (OSM) et l'IG d'autorité et les **indicateurs intrinsèques** en se basant sur la géométrie des objets géographiques issus d'OSM. Deuxièmement, son approche consiste à appliquer des méthodes d'apprentissage pour **estimer les indicateurs extrinsèques à partir des indicateurs intrinsèques**. Le modèle exploré s'appuie sur une régression linéaire multi-variée et utilise les indicateurs intrinsèques comme variables explicatives. La validation du modèle est réalisée en utilisant une approche mixte basée sur la validation croisée 2-fold et bootstrap. Les premiers résultats obtenus sur le thème bâti en zone urbaine sont encourageants et montrent que l'exactitude de position exprimée à travers l'écart surfacique ou l'écart de distance ainsi que l'exactitude de forme exprimée à travers l'écart angulaire et l'écart radial peuvent être expliquées par des indicateurs intrinsèques tels que la granularité, la compacité, la taille, les longueurs minimale, maximale, moyenne, etc. D'autres recherches sont nécessaires pour conclure sur la transférabilité de la méthode pour d'autres types de zone ou d'autres thèmes.

Définition d'une méthode permettant le passage du modèle OSM vers un modèle de référence. Ce travail de recherche qui n'est pas encore abordé dans la thèse d'Ibrahim Maidaneh Abdi pourrait s'appuyer sur les travaux de transformation de schémas dans un contexte utilisateur (Balley, 2007) et permettrait de vérifier des contraintes d'intégrité de la base de données géographiques afin de produire une base de données géographiques conforme aux spécifications définies et qui conserverait les liens avec l'IGV d'origine pour faciliter la mise à jour.

4. Vers la constitution d'un référentiel métier multi-sources

Contrairement à la constitution d'un référentiel à partir d'une source d'IGV, l'objectif dans le cadre du projet Choucas est de construire un référentiel métier en prenant en compte des redondances et des complémentarités de différentes sources de données. Afin d'aboutir à un tel référentiel métier, notre démarche est de s'appuyer sur la modélisation des connaissances à travers les ontologies définies ci-dessus. Plusieurs verrous restent à lever :

- Appariement multi-sources. La plupart des approches d'appariement de données existantes permettent l'appariement de deux sources, ce qui dans le cas de plusieurs sources nécessite de réaliser plusieurs appariements, puis un processus de post-traitement pour établir les liens multi-sources. Ainsi, un premier défi est de pouvoir appairier plusieurs sources d'information géographique en même temps. Cela nécessite donc de définir des stratégies pour la recherche des candidats et la prise de décision. Il ne suffit donc plus seulement de gérer la cardinalité des liens $n:m$, une difficulté en soi, mais de définir des liens d'appariement multi-sources et multi-niveaux.
- Intégration multi-sources proprement dite. Une approche à suivre pourrait être celle proposée par Mustière (2014) qui suggère de proposer un processus complet d'intégration en s'appuyant sur l'appariement de schémas, l'appariement de données, l'analyse des objets géographiques appariés et qui nécessiterait un processus itératif et répétitif. Deux types d'intégration pourraient être envisageables : fusion ou multi-niveaux. Le premier nécessite la définition d'un modèle unique, tandis que le deuxième nécessite de définir les liens entre les différents modèles. Dans le cadre d'un post-doctorat et avec l'encadrement de ma collègue Marie-Dominique Van Damme, nous envisageons d'explorer l'intégration à travers une approche de fusion basée sur les bases de données graphes.
- Identification des relations entre les objets géographiques de différents types. Il s'agit ici de définir les relations entre les objets de repère, les itinéraires et les objets cartographiques afin de pouvoir faciliter l'interrogation des objets géographiques nécessaires à la localisation de victimes en montagne. Les verrous ici sont de pouvoir identifier les relations de localisation entre des objets géographiques de types différents (*e.g.* chalet à droite de l'itinéraire) au niveau du modèle puis de les générer par des mécanismes de raisonnement au niveau des instances. Nous explorons une démarche basée sur les différentes ontologies.

5. Raisonnement spatial flou pour la localisation des victimes en montagne

Une autre thématique de recherche que j'ai abordée dans le cadre de l'encadrement de la thèse de Mattia Bunel est le raisonnement spatial flou pour la localisation des victimes en montagne. Plusieurs objectifs scientifiques sont abordés dans ce travail : la modélisation d'un indice de localisation, le passage d'un référentiel indirect à un référentiel direct, l'agrégation des zones compatibles pour définir des zones de localisation probable où peut se situer la victime et la qualification de cette dernière. Ce raisonnement s'appuie sur les ontologies des objets de référence, cartographiques et itinéraires et des relations de localisation d'une part, et sur des modèles flous basés sur la théorie des sous-ensembles flous, d'autre part. Une attention particulière est accordée à la prise en compte de l'imprécision et de l'incertitude au niveau de l'indice et des concepts modélisant les indices en s'appuyant sur la théorie des fonctions de croyance ou la théorie de possibilités.

6. Vers une aide pour évaluer la qualité de l'information géographique volontaire

Il n'est sans doute plus besoin de justifier le besoin de qualification de l'IG. Que l'IG soit volontaire, d'autorité, ou dérivée à travers des approches d'intégration multi-sources ou mono-source, nous avons besoin d'une part d'évaluer sa qualité interne, de la comparer à d'autres sources d'IG le cas échéant, et de la communiquer aux utilisateurs, et d'autre part d'aider l'utilisateur à évaluer l'utilisabilité d'un jeu de données pour répondre à ses besoins.

Malgré le fait que de nombreux défis restent à relever en lien avec la qualité interne et externe, un grand progrès en recherche a été fait ces dix dernières années. En revanche, il manque à mon sens une plateforme proposant des méthodes, des outils et des benchmarks pour évaluer la qualité. *Pour les questions de benchmarking*, dans le cadre du projet LandSense, nous avons commencé à travailler avec mes collègues de l'Université de Nottingham (Giles Foody) et l'université de Heidelberg (Michael Schultz) pour proposer des benchmarks qui permettront de mesurer l'écart entre un jeu de données d'occupation du sol produit à partir d'une démarche collaborative et un jeu de données faisant autorité ou ayant bénéficié de contrôles qualité.

Une telle plateforme serait bénéfique à la fois aux chercheurs qui travaillent sur la qualité de l'IG, en termes de reproductibilité, aux chercheurs dont l'objet de recherche principal n'est pas la qualité de l'IG mais pour qui cette dernière est une étape primordiale, mais également aux utilisateurs qui auraient besoin de qualifier l'IG. Une telle initiative pourrait être réalisée au sein de la communauté de géomatique française à travers le groupe MAGIS et en particulier via l'*Action Prospective Information Géographique Volontaire et Crowdsourcing pour la connaissance de l'espace géographique*²⁶.

²⁶ <http://gdr-magis.imag.fr/actions-prospectives/information-geographique-volontaire-et-crowdsourcing-pour-la-connaissance-de-lespace-geographique/>

Références

- Acheson, E., De Sabbata, S., S Purves, R., 2017. A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers Environment and Urban Systems* 64, 309–320. <https://doi.org/10.1016/j.compenvurbsys.2017.03.007>
- Al-Bakri, M., Fairbairn, D., 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *International Journal of Geographical Information Science* 26, 1437–1456. <https://doi.org/10.1080/13658816.2011.636012>
- Alvares, L.A., Fernandes, J.A., Macedo, D., Bogorny, V., Moelans, B., Kuijpers, B., Vaisman, A., 2007. A model for enriching trajectories with semantic geographical information., in: 15th Annual ACM International Symposium on Advances in GIS.
- Antoniou, V., Morley, J., Haklay, M., 2009. The role of user generated spatial content in mapping agencies, in: *Proceedings of GISRUK Conference*.
- Antoniou, V., Schlieder, C., 2018. Addressing Uneven Participation Patterns in VGI Through Gamification Mechanisms, in: Ahlqvist, O., Schlieder, C. (Eds.), *Geogames and Geoplay, Advances in Geographic Information Science*. Springer International Publishing, Cham, pp. 91–110. https://doi.org/10.1007/978-3-319-22774-0_5
- Antoniou, V., Skopeliti, A., 2017. The Impact of the Contribution Micro-environment on Data Quality: The Case of OSM, in: *Mapping and the Citizen Sensor*. London: Ubiquity Press, pp. 165–196.
- Antoniou, V., Skopeliti, A., 2015. Measures and indicators of VGI quality: AN OVERVIEW. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5, 345–351. <https://doi.org/10.5194/isprsannals-ii-3-w5-345-2015>
- Antoniou, V., Touya, G., Olteanu-Raimond, A.-M., 2016. Quality analysis of the Parisian OSM toponyms evolution, in: *European Handbook of Crowdsourced Geographic Information*. ubiquity press London, London.
- Arnold, S., Kosztra, B., Banko, G., Milenov, P., Smith, G., Hazeu, G., 2016. Content Documentation of the EAGLE Concept (Livable No. 2.3.1).
- Ballatore, A., Zipf, A., 2015. A Conceptual Quality Framework for Volunteered Geographic Information, in: Fabrikant, S.I., Raubal, M., Bertolotto, M., Davies, C., Friendschuh, S., Bell, S. (Eds.), *Spatial Information Theory, Lecture Notes in Computer Science*. Springer International Publishing, pp. 89–107. https://doi.org/10.1007/978-3-319-23374-1_5
- Balley, S., 2007. Aide à la restructuration de données géographiques sur le Web - Vers la diffusion à la carte d'information géographique. Université Paris-Est.
- Barron, C., Neis, P., Zipf, A., 2014. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS* 18, 877–895. <https://doi.org/10.1111/tgis.12073>
- Basiouka, S., Potsiou, C., 2012. VGI in Cadastre: a Greek experiment to investigate the potential of crowd sourcing techniques in Cadastral Mapping. *Survey Review* 44, 153–161. <https://doi.org/10.1179/1752270611Y.0000000037>
- Bateman, J.A., Hois, J., Ross, R., Tenbrink, T., 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence* 174, 1027–1071. <https://doi.org/10.1016/j.artint.2010.05.008>
- Bégin, D., 2014. Towards integrating VGI and national mapping agency operations - a Canadian case study. Presented at the Workshop on Role of Volunteered Geographic Information in Advancing Science: Quality and Credibility, Columbus (OH), USA.
- Bégin, D., Devillers, R., Roche, S., 2018. The life cycle of contributors in collaborative online communities -the case of OpenStreetMap. *International Journal of Geographical Information Science* 32, 1611–1630. <https://doi.org/10.1080/13658816.2018.1458312>
- Bishr, M., Kuhn, W., 2007. Geospatial Information Bottom-Up: A Matter of Trust and Semantics, in: Fabrikant, S.I., Wachowicz, M. (Eds.), *The European Information Society: Leading the Way*

- with Geo-Information. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 365–387. https://doi.org/10.1007/978-3-540-72385-1_22
- Bol, D., Grus, M., Laakso, M., 2016. Crowdsourcing and VGI in National Mapping Agency's Data Collection 6.
- Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, 977–984. <https://doi.org/10.1525/bio.2009.59.11.9>
- Bordogna, G., Frigerio, L., Kliment, T., Brivio, P., Hossard, L., Manfron, G., Sterlacchini, S., 2016. "Contextualized VGI" Creation and Management to Cope with Uncertainty and Imprecision. *ISPRS International Journal of Geo-Information* 5, 234. <https://doi.org/10.3390/ijgi5120234>
- Bovolo, F., Bruzzone, L., 2007. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Transactions on Geoscience and Remote Sensing* 45, 218–236.
- Brando, C., Bucher, B., Abadie, N., 2011. Specifications for User Generated Spatial Content, in: Geertman, S., Reinhardt, W., Toppen, F. (Eds.), *Advancing Geoinformation Science for a Changing World, Lecture Notes in Geoinformation and Cartography*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 479–495. https://doi.org/10.1007/978-3-642-19789-5_24
- Bruns, A., 2008. Blogs, Wikipedia, Second life, and Beyond : from production to produsage. Peter Lang.
- Buard, E., 2011. Pratiques spatiales des populations animales: analyses par les trajectoires, in: *Actes de Dixièmes Rencontres de Théo Quant*.
- Bucher, B., Falquet, G., Metral, C., Lemmens, R., 2016. Enhancing the management of quality of VGI: contributions from context and task modelling, in: *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press.
- Budhathoki, N.R., Haythornthwaite, C., 2013. Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist* 57, 548–575. <https://doi.org/10.1177/0002764212469364>
- Bunel, M., Duchêne, C., Olteanu-Raimond, A.-M., Villanova-Oliver, M., Bonhoure, G., Jouan, T., 2019a. An approach for identifying and analysing reference features and spatial relations used in mountain emergency calls. *Proceedings of the ICA* 2, 1–8. <https://doi.org/10.5194/ica-proc-2-12-2019>
- Bunel, M., Olteanu-Raimond, A.-M., Duchêne, C., 2019b. Objets et relations spatiales composites et prise en compte du vague pour interpréter un référencement spatial indirect. *Revue Internationale de Géomatique* 29, 81–106. <https://doi.org/10.3166/rig.2019.00075>
- Büttner, G., 2014. CORINE Land Cover and Land Cover Change Products, in: Manakos, I., Braun, M. (Eds.), *Land Use and Land Cover Mapping in Europe, Remote Sensing and Digital Image Processing*. Springer Netherlands, Dordrecht, pp. 55–74. https://doi.org/10.1007/978-94-007-7969-3_5
- Cain, J.W., Krausman, P.R., Jansen, B.D., Morgart, J.R., 2005. Influence of topography and GPS fix interval on GPS collar performance. *Wildlife Society Bulletin* 33, 926–934. [https://doi.org/10.2193/0091-7648\(2005\)33%5B926:iotagf%5D2.0.co;2](https://doi.org/10.2193/0091-7648(2005)33%5B926:iotagf%5D2.0.co;2)
- Campagna, M., Floris, R., Massa, P., Girsheva, A., Ivanov, K., 2015. The Role of Social Media Geographic Information (SMGI) in Spatial Planning, in: Geertman, S., Ferreira, J., Goodspeed, R., Stillwell, J. (Eds.), *Planning Support Systems and Smart Cities, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham, pp. 41–60. https://doi.org/10.1007/978-3-319-18368-8_3
- Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F., Purves, R. (Eds.), 2016. *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press. <https://doi.org/10.5334/bax.h>
- Chan, K.S., Godby, R., Mestelman, S., Andrew Muller, R., 2002. Crowding-out voluntary contributions to public goods. *Journal of Economic Behavior & Organization* 48, 305–317. [https://doi.org/10.1016/S0167-2681\(01\)00232-3](https://doi.org/10.1016/S0167-2681(01)00232-3)

- Christophe, S., 2017. Personnalisation des représentations cartographiques en géovisualisation : couleurs et styles (HDR). Université Paris-Est.
- Cohen, W.W., 1995. Fast Effective Rule Induction, in: In Proceedings of the Twelfth International Conference on Machine Learning. Morgan Kaufmann, pp. 115–123.
- Comber, A., Fisher, P., Wadsworth, R., 2004. Integrating land-cover data with different ontologies: identifying change from inconsistency. *International Journal of Geographical Information Science* 18, 691–708. <https://doi.org/10.1080/13658810410001705316>
- Comber, A.J., Law, A.N.R., Lishman, J.R., 2004. A comparison of Bayes', Dempster-Shafer and Endorsement theories for managing knowledge uncertainty in the context of land cover monitoring. *Computers, Environment and Urban Systems* 28, 311–327. [https://doi.org/10.1016/S0198-9715\(03\)00013-9](https://doi.org/10.1016/S0198-9715(03)00013-9)
- Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J., 2013. #Earthquake: Twitter as a Distributed Sensor System: #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS* 17, 124–147. <https://doi.org/10.1111/j.1467-9671.2012.01359.x>
- DeCesare, N.J., Kolbe, J.A., Squires, J.R., 2005. Effect of forest canopy on GPS-based movement data 33, 935–941.
- DeCesare, Nicholas J., Squires, J.R., Kolbe, J.A., 2005. Effect of forest canopy on GPS-based movement data. *Wildlife Society Bulletin* 33, 935–941. [https://doi.org/10.2193/0091-7648\(2005\)33%5B935:eofcog%5D2.0.co;2](https://doi.org/10.2193/0091-7648(2005)33%5B935:eofcog%5D2.0.co;2)
- Denis, M., 1997. The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive* 16, 409–458.
- D'Eon, R.G., Delparte, D., 2005. Effects of radio-collar position and orientation on GPS radio-collar performance, and the implications of PDOP in data screening. *Journal of Applied Ecology* 42, 383–388. <https://doi.org/10.1111/j.1365-2664.2005.01010.x>
- Derungs, C., S. Purves, R., 2013. From text to landscape: Locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus. *International Journal of Geographical Information Science* 28, 1272–1293. <https://doi.org/10.1080/13658816.2013.772184>
- Devillers, R., Desjardin, E., De Runz, C., 2019. Concepts et terminologies de l'imperfection de l'information géographique, in: *L'imperfection Des Données Géographiques* 1. pp. 13–27.
- Devillers, R., Jeansoulin, R., 2005. *Information Géographique et Aménagement du Territoire*.
- Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., Shi, W., 2010. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities: Thirty Years of Research on Spatial Data Quality. *Transactions in GIS* 14, 387–400. <https://doi.org/10.1111/j.1467-9671.2010.01212.x>
- Domenech, E., Mallet, C., 2014. Change Detection in High-Resolution Land Use/Land Cover Geodatabases (at Object Level). *EuroSDR Official Publication* 64.
- Dominguès, C., Corby, O., Soualah-Alila, F., 2012. Reasonner sur une ontologie cartographique pour concevoir des légendes de cartes. Presented at the 12e Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances, Editions Hermann, Bordeaux, France, pp. 309–314.
- EEA, 2018. Mapping Guide for a European Urban Atlas. 2011 [WWW Document]. Copernicus Land Monitoring Service - Urban Atlas. URL <https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-urban-atlas> (accessed 5.6.20).
- Egenhofer, M.J., Mark, D.M., 1995. Naive Geography, in: Frank, A.U., Kuhn, W. (Eds.), *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–15.
- Egorova, E., Tenbrink, T., Purves, R.S., 2015. Where Snow is a Landmark: Route Direction Elements in Alpine Contexts, in: Fabrikant, S.I., Raubal, M., Bertolotto, M., Davies, C., Freundschuh, S., Bell, S. (Eds.), *Spatial Information Theory, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 175–195. https://doi.org/10.1007/978-3-319-23374-1_9
- Eliasson, M., 2014. A Kalman filter approach to reduce position error for pedestrian applications in areas of bad GPS reception (Master's Thesis). Universitet Umea.

- Elwood, S., Goodchild, M.F., Sui, D.Z., 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers* 102, 571–590. <https://doi.org/10.1080/00045608.2011.595657>
- Estellés-Arolas, E., González-Ladrón-de-Guevara, F., 2012. Towards an integrated crowdsourcing definition. *Journal of Information Science* 38, 189–200. <https://doi.org/10.1177/0165551512437638>
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96*. AAAI Press, Portland, Oregon, pp. 226–231.
- Estima, J., Painho, M., 2013. Exploratory analysis of OpenStreetMap for land use classification, in: *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information - GEOCROWD '13*. Presented at the the Second ACM SIGSPATIAL International Workshop, ACM Press, Orlando, Florida, pp. 39–46. <https://doi.org/10.1145/2534732.2534734>
- Etienne, L., 2011. Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction a la détection de comportements inhabituels. Application au trafic maritime. (PhD Thesis). Université de Bretagne occidentale, Brest.
- Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science* 28, 700–719. <https://doi.org/10.1080/13658816.2013.867495>
- Flanagin, A., Metzger, M., 2008. The credibility of volunteered geographic information. *GeoJournal* 72, 137–148. <https://doi.org/10.1007/s10708-008-9188-y>
- Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 378–382.
- Follin, J.-M., Girres, J.-F., Olteanu-Raimond, A.-M., Sheeren, D., 2019. Origine de l'imperfection de l'information géographique, in: *L'imperfection Des Données Géographiques 1*. ISTE Editions, pp. 29–51.
- Fonte, C., Minghini, M., Patriarca, J., Antoniou, V., See, L., Skopeliti, A., 2017a. Generating Up-to-Date and Detailed Land Use and Land Cover Maps Using OpenStreetMap and GlobeLand30. *ISPRS International Journal of Geo-Information* 6, 125. <https://doi.org/10.3390/ijgi6040125>
- Fonte, C., Patriarca, J., Minghini, M., Antoniou, V., See, L., Brovelli, M., 2017b. Using openstreetmap to create land use and land cover maps: Development of an application, in: *Volunteered Geographic Information and the Future of Geospatial Data*. pp. 113–137.
- Fonte, C. C., Bastin, L., Foody, G., Kellenberger, T., Kerle, N., Mooney, P., Olteanu-Raimond, A.-M., See, L., 2015. VGI QUALITY CONTROL. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W5*, 317–324. <https://doi.org/10.5194/isprsannals-II-3-W5-317-2015>
- Fonte, Cidalia C., Bastin, L., See, L., Foody, G., Lupia, F., 2015. Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science* 29, 1269–1291. <https://doi.org/10.1080/13658816.2015.1018266>
- Foody, G.M., 2009. Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing* 30, 5273–5291. <https://doi.org/10.1080/01431160903130937>
- Foody, G.M., Boyd, D.S., 2012. Using volunteered data in land cover map validation: Mapping tropical forests across West Africa, in: *2012 IEEE International Geoscience and Remote Sensing Symposium*. Presented at the IGARSS 2012 - 2012 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Munich, Germany, pp. 6207–6208. <https://doi.org/10.1109/IGARSS.2012.6352675>
- Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing* 1, 345–354. <https://doi.org/10.3390/rs1030345>

- Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., Obersteiner, M., 2012. Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software* 31, 110–123. <https://doi.org/10.1016/j.envsoft.2011.11.015>
- Fritz, S., See, L., Brovelli, M., 2017. Motivating and Sustaining Participation in VGI, in: *Mapping and the Citizen Sensor*. Ubiquity Press, pp. 93–117.
- Girres, J.-F., 2012. *Modèle d'estimation de l'imprécision des mesures géométriques de données géographiques - Application aux mesures de longueur et de surface (Thèse)*. Université Paris-Est.
- Girres, J.-F., Touya, G., 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS* 14, 435–459. <https://doi.org/10.1111/j.1467-9671.2010.01203.x>
- Gómez-Barrón, J.-P., Manso-Callejo, M.-Á., Alcarria, R., Iturrioz, T., 2016. Volunteered Geographic Information System Design: Project and Participation Guidelines. *ISPRS International Journal of Geo-Information* 5, 108. <https://doi.org/10.3390/ijgi5070108>
- Gomez-Gil, J., Ruiz-Gonzalez, R., Alonso-Garcia, S., Gomez-Gil, F., 2013. A Kalman Filter Implementation for Precision Improvement in Low-Cost GPS Positioning of Tractors. *Sensors* 13, 15307–15323. <https://doi.org/10.3390/s131115307>
- 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., Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth* 3, 231–241. <https://doi.org/10.1080/17538941003759255>
- Goodchild, M.F., Li, L., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics* 1, 110–120. <https://doi.org/10.1016/j.spasta.2012.03.002>
- Gressin, A., Vincent, N., Mallet, C., Paparoditis, N., 2014. A unified framework for land-cover database update and enrichment using satellite imagery, in: *Proc. of the IEEE International Conference on Image Processing (ICIP)*. Paris, France.
- Grus, M., Laakso, M., 2016. Experiences : Introducing data collection in schools with a mobile application #hylo.
- Haklay, M., 2010. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design* 37, 682–703. <https://doi.org/10.1068/b35097>
- Haklay, M., Antoniou, V., Basiouka, S., Soden Rober, Mooney, P., 2014. Crowd-sourced geographic information use in government (Working paper No. 92239). Washington, DC : World Bank Group.
- Halkosaari, H.-M., Rönneberg, M., Laakso, M., Kettunen, P., Oksanen, J., Sarjakoski, T., 2017. Concept Design of #hylo—Geosocial Network for Sharing Hyperlocal Information on a Map, in: Gartner, G., Huang, H. (Eds.), *Progress in Location-Based Services 2016, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham, pp. 309–327. https://doi.org/10.1007/978-3-319-47289-8_16
- Hartter, J., Ryan, S.J., MacKenzie, C.A., Parker, J.N., Strasser, C.A., 2013. Spatially Explicit Data: Stewardship and Ethical Challenges in Science. *PLoS Biology* 11, e1001634. <https://doi.org/10.1371/journal.pbio.1001634>
- Hashemi, P., Abbaspour, R.A., 2015. Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept, in: *Openstreetmap in Giscience*. Springer, pp. 19–36.
- Hayat, F., 2019. *Production des biens communs numériques et usages cartographiques (Mémoire de thèse)*. Université Sorbonne Paris Cité Préparée à l'Université Paris Diderot.
- Heipke, C., 2010. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 550–557. <https://doi.org/10.1016/j.isprsjprs.2010.06.005>
- Hill, L., Zheng, Q., 1999. Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers.

- Presented at the Proceedings of the 62nd Annual Meeting of the American Society for Information Science, Washington, DC, USA, pp. 57–69.
- Hollenstein, L., Purves, R., 2010. Exploring place through user-generated content: Using flickr tags to describe city cores. *Journal of Spatial Information Science* 21–48.
- Houiller, F., Merilhou-Goudard, J.-B., 2016. Les Sciences participatives en France, État des lieux, bonnes pratiques & recommandations.
- INSPIRE, 2013. Data Specification on Land use – Draft Technical Guidelines (Spécifications).
- Iovan, C., Olteanu-Raimond, A.-M., Couronné, T., Smoreda, Z., 2013. Moving and Calling: Mobile Phone Data Quality Measurements and Spatiotemporal Uncertainty in Human Mobility Studies, in: Vandenbroucke, D., Bucher, B., Crompvoets, J. (Eds.), *Geographic Information Science at the Heart of Europe, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, pp. 247–265. https://doi.org/10.1007/978-3-319-00615-4_14
- Ivanovic, S., 2018. Quality based approach for updating geographic authoritative datasets from crowdsourced GPS traces (Thèse). Université Paris-Est, Paris, France.
- Ivanovic, S., Olteanu-Raimond, A.-M., Mustière, S., Devogele, T., 2019. A Filtering-Based Approach for Improving Crowdsourced GNSS Traces in a Data Update Context. *ISPRS International Journal of Geo-Information* 8, 380. <https://doi.org/10.3390/ijgi8090380>
- Ivanovic, S.S., Olteanu-Raimond, A.-M., Mustière, S., Devogele, T., 2019. Potential of Crowdsourced Traces for Detecting Updates in Authoritative Geographic Data, in: Kyriakidis, P., Hadjimitsis, D., Skarlatos, D., Mansourian, A. (Eds.), *Geospatial Technologies for Local and Regional Development, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham, pp. 205–221. https://doi.org/10.1007/978-3-030-14745-7_12
- Japkowicz, N., 2000. Learning from Imbalanced Data Sets: A Comparison of Various Strategies. AAAI.
- Johnson, P.A., Sieber, R.E., 2013. Situating the Adoption of VGI by Government, in: Sui, D., Elwood, S., Goodchild, M. (Eds.), *Crowdsourcing Geographic Knowledge*. Springer Netherlands, Dordrecht, pp. 65–81. https://doi.org/10.1007/978-94-007-4587-2_5
- Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., Zipf, A., 2013. Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science* 27, 2264–2278. <https://doi.org/10.1080/13658816.2013.800871>
- Jokar Arsanjani, J., Mooney, P., Zipf, A., Schauss, A., 2015a. Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets, in: *OpenStreetMap in GIScience*. Springer International Publishing, Cham, pp. 37–58.
- Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M. (Eds.), 2015b. *OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-14280-7>
- Joliveau, T., 2011. Le géoweb, un nouveau défi pour les bases de données géographiques. *Espace géographique* 40, 154. <https://doi.org/10.3917/eg.402.0154>
- Jolivet, L., Olteanu-Raimond, A.-M., 2017. Crowd and Community Sourced Data Quality Assessment, in: Peterson, M.P. (Ed.), *Advances in Cartography and GIScience*. Springer International Publishing, Cham, pp. 47–60. https://doi.org/10.1007/978-3-319-57336-6_4
- Kádár, B., Gede, M., 2013. Where Do Tourists Go? Visualizing and Analysing the Spatial Distribution of Geotagged Photography. *Cartographica: The International Journal for Geographic Information and Geovisualization* 48, 78–88. <https://doi.org/10.3138/carto.48.2.1839>
- Karimipour, F., Azari, O., 2015. Citizens as Expert Sensors: One Step Up on the VGI Ladder, in: Gartner, G., Huang, H. (Eds.), *Progress in Location-Based Services 2014, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham, pp. 213–222. https://doi.org/10.1007/978-3-319-11879-6_15
- Keßler, C., de Groot, R.T.A., 2013. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap, in: *Geographic Information Science at the Heart of Europe*. Springer International Publishing, Cham, pp. 21–37.
- Krumm, J., Davies, N., Narayanaswami, C., 2008. User-Generated Content. *IEEE Pervasive Computing* 7, 10–11. <https://doi.org/10.1109/mprv.2008.85>

- Larousserie, D., 2013. Un village landais reprend possession de ses données. *Le Monde*.
- Laso Bayas, J., See, L., Fritz, S., Sturn, T., Perger, C., Dürauer, M., Karner, M., Moorthy, I., Schepaschenko, D., Domian, D., McCallum, I., 2016. Crowdsourcing In-Situ Data on Land Cover and Land Use Using Gamification and Mobile Technology. *Remote Sensing* 8, 905. <https://doi.org/10.3390/rs8110905>
- Le Bris, A., Chehata, N., 2011. Change detection in a topographic building database using submetric satellite images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(3/W22), 25–30.
- Lewis, J.S., Rachlow, J.L., Garton, E.O., Vierling, L.A., 2007. Effects of habitat on GPS collar performance: using data screening to reduce location error. *Journal of Applied Ecology* 44, 663–671. <https://doi.org/10.1111/j.1365-2664.2007.01286.x>
- Liu, C., Xiong, L., Hu, X., Shan, J., 2015. A Progressive Buffering Method for Road Map Update Using OpenStreetMap Data. *ISPRS International Journal of Geo-Information* 4, 1246–1264. <https://doi.org/10.3390/ijgi4031246>
- Liu, L., Olteanu-Raimond, A.-M., Jolivet, L., Le Bris, A., 2019. Operational workflows for integration of citizen-observed data in authoritative systems (Livable No. D5.5).
- Liu, S., Bruzzone, L., Bovolo, F., Du, P., 2015. Hierarchical unsupervised change detection in multitemporal hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing* 53, 244–260.
- Lynch, K., 1960. *The Image of the City*. MIT Press.
- Matheus, A., Batič, M., 2016. LandSense Engagement Platform II (Livable No. D3.6).
- Medad, A., Gaio, M., Le Nir, Y., 2018. Appariement automatique de données hétérogènes: textes, traces GPS et ressources géographiques. Presented at the Sagéo.
- Minghini, M., Antoniou, V., Fonte, C.C., Estima, J., Olteanu-Raimond, A.-M., See, L., Mari Laakso, Skopeliti, A., Mooney, P., Arsanjani, J.J., Lupio, F., 2017. The Relevance of Protocols for VGI Collection, in: *Mapping and the Citizen Sensor*. pp. 223–247.
- Mohammadi, N., Malek, M., 2015. Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data. *Journal of Spatial Science* 60, 119–135. <https://doi.org/10.1080/14498596.2014.927337>
- Moncla, L., Renteria-Agualimpia, W., Nogueras-Iso, J., Gaio, M., 2014. Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus, in: *ACM (Ed.), ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2014)*. Dallas, Texas, United States. <https://doi.org/10.1145/2666310.2666386>
- Mooney, P., Morley, J., 2014. *Crowdsourcing in National Mapping* (No. 64). Published by EuroESDR: Official Publication.
- Mullen, W.F., Jackson, S.P., Croitoru, A., Crooks, A., Stefanidis, A., Agouris, P., 2015. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *GeoJournal* 80, 587–605. <https://doi.org/10.1007/s10708-014-9564-8>
- Mustière, S., 2014. *Intégration de données géographiques*. Université Paris-Est.
- Mustière, S., 2001. *Apprentissage supervisé pour la généralisation cartographique (Thèse)*. Université Paris-Est.
- Mustière, S., Abadie, N., Aussenac-Gilles, N., Bessagnet, M.-N., Kamiel, M., Kergosien, E., Reynaud, C., Safar, B., Sallaberry, C., 2011a. Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique. *Revue internationale de géomatique* 21, 155–179. <https://doi.org/10.3166/rig.21.155-179>
- Mustière, S., Abadie, N., Aussenac-Gilles, N., Bessagnet, M.-N., Kamiel, M., Kergosien, E., Reynaud, C., Safar, B., Sallaberry, C., 2011b. Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique. *Revue internationale de géomatique* 21, 155–179. <https://doi.org/10.3166/rig.21.155-179>
- Mustière, S., Devogele, T., 2008. Matching Networks with Different Levels of Detail. *GeoInformatica* 12, 435–453. <https://doi.org/10.1007/s10707-007-0040-1>

- Neis, P., Zielstra, D., Zipf, A., 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet* 5, 282–300. <https://doi.org/10.3390/fi5020282>
- Neis, P., Zielstra, D., Zipf, A., 2011. The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet* 4, 1–21. <https://doi.org/10.3390/fi4010001>
- Newson, P., Krumm, J., 2009. Hidden Markov Map Matching Through Noise and Sparseness, in: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*. ACM, New York, NY, USA, pp. 336–343. <https://doi.org/10.1145/1653771.1653818>
- Nielsen, J., 2016. The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities. Nielsen Norman Group.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>
- Olteanu, A.M., 2008. Fusion de connaissances imparfaites pour l'appariement de données géographiques : proposition d'une approche s'appuyant sur la théorie des fonctions de croyance (PhD Thesis). Université Paris-Est.
- Olteanu-Raimond, A.-M., Couronné, T., Fen-Chong, J., Smoreda, Z., 2011. Modélisation des trajectoires spatio-temporelles issues des traces numériques de téléphones portables. *Le Paris des visiteurs, qu'en disent les téléphones mobiles?*
- Olteanu-Raimond, A.-M., Hart, G., Foody, G.M., Touya, G., Kellenberger, T., Demetriou, D., 2017a. The Scale of VGI in Map Production: A Perspective on European National Mapping Agencies: VGI and the National Mapping Agencies. *Transactions in GIS* 21, 74–90. <https://doi.org/10.1111/tgis.12189>
- Olteanu-Raimond, A.-M., Jolivet, L., Van Damme, M.-D., Royer, T., Fraval, L., See, L., Sturn, T., Karner, M., Moorthy, I., Fritz, S., 2018. An Experimental Framework for Integrating Citizen and Community Science into Land Cover, Land Use, and Land Change Detection Processes in a National Mapping Agency. *Land* 7, 103. <https://doi.org/10.3390/land7030103>
- Olteanu-Raimond, A.-M., Laakso, Mari, Antoniou, Vyron, Fonte, C.C., Fonseca, Alexandra, Magdalena Grus, Harding, Jenny, Kellenberger, Tobias, Minghini, Marco, Andriani Skopeliti, 2017b. VGI in National Mapping Agencies: Experiences and Recommendations, in: *Mapping and the Citizen Sensor*. Ubiquity Press, pp. 299–326.
- Olteanu-Raimond, Ana-Maria, Mustière, S., Bunel, M., Domingues, C., Duchêne, C., Jolivet, L., Van Damme, M.-D., 2020. Ontologie d'objets de repères Choucas (Livable projet Choucas).
- Olteanu-Raimond, A.-M., Mustière, S., Ruas, A., 2015. Knowledge formalization for vector data matching using belief theory. *Journal of Spatial Information Science*. <https://doi.org/10.5311/JOSIS.2015.10.194>
- Olteanu-Raimond, A.-M., See, L., Schultz, M., Foody, G., Riffler, M., Gasber, T., Jolivet, L., le Bris, A., Meneroux, Y., Liu, L., Poupée, M., Gombert, M., 2020. Use of Automated Change Detection and VGI Sources for Identifying and Validating Urban Land Use Change. *Remote Sensing* 12, 1186. <https://doi.org/10.3390/rs12071186>
- Perret, J., Gribaudi, M., Barthelemy, M., 2015. Roads and cities of 18th century France. *Scientific Data* 2. <https://doi.org/10.1038/sdata.2015.48>
- Pilz, D., Ballard, H.L., Jones, E.T., 2006. Broadening Participation in Biological Monitoring: Guidelines for Scientists and Managers Institute for Culture and Ecology (General Technical Report PNW-GTR-680).
- Pocock, M.J.O., Chapman, D.S., Sheppard, L.J., Roy, H.E., 2014. Choosing and using citizen science: a guide to when and how to use citizen science to monitor biodiversity and the environment.
- Rice, M.T., Jacobson, R.D., Caldwell, D.R., McDermott, S.D., Paez, F.I., Aburizaiza, A.O., Curtin, K.M., Stefanidis, A., Qin, H., 2013. Crowdsourcing techniques for augmenting traditional

- accessibility maps with transitory obstacle information. *Cartography and Geographic Information Science* 40, 210–219. <https://doi.org/10.1080/15230406.2013.799737>
- Rönneberg, M., Laakso, M., Sarjakoski, T., 2019. Map Gretel: social map service supporting a national mapping agency in data collection. *Journal of Geographical Systems* 21, 43–59. <https://doi.org/10.1007/s10109-018-0288-z>
- Roszkowska, E., 2013. Rank ordering criteria weighting methods - a comparative overview. *Optimum. Studia Ekonomiczne* 14–33.
- Ruas, A., 2002. Généralisation et représentation multiple. Lavoisier.
- Ruas, A., Plumejeaud, C., Nahassia, L., Grosso, E., Olteanu, A.-M., Costes, B., Vouloir, M.-C., Motte, C., 2014. GéoPeuple: The Creation and the Analysis of Topographic and Demographic Data Over 200 Years, in: Buchroithner, M., Prechtel, N., Burghardt, D. (Eds.), *Cartography from Pole to Pole, Lecture Notes in Geoinformation and Cartography*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–17. https://doi.org/10.1007/978-3-642-32618-9_1
- Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake shakes Twitter users: real-time event detection by social sensors, in: *Proceedings of the 19th International Conference on World Wide Web - WWW '10*. Presented at the the 19th international conference, ACM Press, Raleigh, North Carolina, USA, p. 851. <https://doi.org/10.1145/1772690.1772777>
- Scassa, T., 2013. Legal issues with volunteered geographic information: Legal issues with volunteered geographic information. *The Canadian Geographer / Le Géographe canadien* 57, 1–10. <https://doi.org/10.1111/j.1541-0064.2012.00444.x>
- Schultz, M., Voss, J., Auer, M., Carter, S., Zipf, A., 2017. Open land cover from OpenStreetMap and remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 63, 206–213. <https://doi.org/10.1016/j.jag.2017.07.014>
- Sedano, E., 2016. 'Sensor'ship and Spatial Data Quality. *Urban Planning* 1, 75. <https://doi.org/10.17645/up.v1i2.608>
- See, L., Laso Bayas, J., Schepaschenko, D., Perger, C., Dresel, C., Maus, V., Salk, C., Weichselbaum, J., Lesiv, M., McCallum, I., Moorthy, I., Fritz, S., 2017. LACO-Wiki: A New Online Land Cover Validation Tool Demonstrated Using GlobeLand30 for Kenya. *Remote Sensing* 9, 754. <https://doi.org/10.3390/rs9070754>
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Póddör, A., Olteanu-Raimond, A.-M., Rutzinger, M., 2016. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information* 5, 55. <https://doi.org/10.3390/ijgi5050055>
- Senaratne, H., Mobasher, A., Ali, A.L., Capineri, C., Haklay, M. (Muki), 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science* 31, 139–167. <https://doi.org/10.1080/13658816.2016.1189556>
- Senaratne, H., Mobasher, A., Ali, A.L., Capineri, C., Haklay, M.M., O. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science* 0, 1–29. <https://doi.org/10.1080/13658816.2016.1189556>
- Sheeren, D., 2005. Compréhension et interprétation des différences de représentation entre les données géographiques (Thèse). Université Paris-Est.
- Shen, N., Yang, J., Yuan, K., Fu, C., Jia, C., 2016. An efficient and privacy-preserving location sharing mechanism. *Computer Standards & Interfaces* 44, 102–109. <https://doi.org/10.1016/j.csi.2015.06.001>
- Sherwin, W.B., Prat i Fornells, N., 2019. The Introduction of Entropy and Information Methods to Ecology by Ramon Margalef. *Entropy* 21, 794. <https://doi.org/10.3390/e21080794>
- Sieber, R., 2006. Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers* 96, 491–507. <https://doi.org/10.1111/j.1467-8306.2006.00702.x>

- Silva, C.N. (Ed.), 2013. Citizen E-Participation in Urban Governance: Crowdsourcing and Collaborative Creativity, *Advances in Electronic Government, Digital Divide, and Regional Development*. IGI Global. <https://doi.org/10.4018/978-1-4666-4169-3>
- Sinha, G., Arundel, S.T., Hahmann, T., Usery, E.L., Stewart, K., Mark, D.M., 2018. The Landform Reference Ontology (LFRO): A Foundation for Exploring Linguistic and Geospatial Conceptualization of Landforms (Short Paper). Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany. <https://doi.org/10.4230/lipics.giscience.2018.59>
- Sitthi, A., Nagai, M., Dailey, M., Ninsawat, S., 2016. Exploring Land Use and Land Cover of Geotagged Social-Sensing Images Using Naive Bayes Classifier. *Sustainability* 8, 921. <https://doi.org/10.3390/su8090921>
- Spécifications BDTOPO, 2019. BDTOPO V3 - Descriptif de contenu (Spécifications No. V0.3). IGN.
- Stefanidis, A., Crooks, A., Radzikowski, J., 2013. Harvesting ambient geospatial information from social media feeds. *GeoJournal* 78, 319–338. <https://doi.org/10.1007/s10708-011-9438-2>
- Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment* 231, 111199. <https://doi.org/10.1016/j.rse.2019.05.018>
- Sui, D.Z., 2007. Volunteered Geographic Information: A tetradic analysis using McLuhan's law of the media. Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, December 13-14, 2007.
- Teulade-Denantes, J., Maudet, A., Duchêne, C., 2015. Routes visualization: Automated placement of multiple route symbols along a physical network infrastructure. *Journal of Spatial Information Science (JoSIS)* 11, 53–79.
- Thierry, B., Chaix, B., Kestens, Y., 2013. Detecting activity locations from raw GPS data: a novel kernel-based algorithm. *International Journal of Health Geographics* 12, 14+. <https://doi.org/10.1186/1476-072x-12-14>
- Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46, 234. <https://doi.org/10.2307/143141>
- Touya, G., 2017. Vers l'automatisation de la production de cartes (HDR). Université Paris Est.
- Touya, G., Antoniou, V., Olteanu-Raimond, A.-M., Van Damme, M.-D., 2017. Assessing Crowdsourced POI Quality: Combining Methods Based on Reference Data, History, and Spatial Relations. *ISPRS International Journal of Geo-Information* 6, 80. <https://doi.org/10.3390/ijgi6030080>
- Touya, G., Brando-Escobar, C., 2013. Detecting Level-of-Detail Inconsistencies in Volunteered Geographic Information Data Sets. *Cartographica: The International Journal for Geographic Information and Geovisualization* 48, 134–143. <https://doi.org/10.3138/carto.48.2.1836>
- Truong, Q.T., 2020. Le vandalisme de l'information géographique volontaire, Analyse exploratoire et proposition d'une méthodologie de détection automatique. Université Paris-Est.
- Truong, Q.T., De Runz, C., Touya, G., 2019. Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *International Journal of Geographical Information Science* 33, 1651–1682.
- Turner, A., 2006. *Introduction to Neogeography*. O'Reilly, Sebastopol, Calif.
- Van Damme, M.-D., Olteanu-Raimond, A.-M., Méneroux, Y., 2019. Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas. *International Journal of Cartography* 5, 195–213. <https://doi.org/10.1080/23729333.2019.1615730>
- Van Exel, M., Dias, E., Fruijt, S., 2010. The impact of crowdsourcing on spatial data quality indicators, in: *In Proceedings of GiScience 2011*.
- van Winden, K., Biljecki, F., van der Spek, S., 2016. Automatic Update of Road Attributes by Mining GPS Tracks. *Transactions in GIS* 20, 664–683. <https://doi.org/10.1111/tgis.12186>
- Viglino, J.-M., 2011. Vers un système collaboratif pour la mise à jour de référentiels géographiques, in: *Proceedings of the 25th International Cartographic Conference*. Presented at the International Cartographic Conference, Paris, France.

- Viry, M., Villanova-Oliver, M., 2019. L'Ontologie d'Alerte Choucas : de la modélisation des connaissances à un outil support d'un raisonnement géovisuel. Presented at the Sagéo, Clermont-Ferrand, France.
- Walter, V., Fritsch, D., 1999. Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science* 13, 445–473. <https://doi.org/10.1080/136588199241157>
- Wang, R.Y., Strong, D.M., 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12, 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- Xie, P., Petovello, M.G., 2015. Measuring GNSS Multipath Distributions in Urban Canyon Environments. *IEEE Transactions on Instrumentation and Measurement* 64, 366–377. <https://doi.org/10.1109/tim.2014.2342452>
- Xu, Y., Chen, Z., Xie, Z., Wu, L., 2017. Quality assessment of building footprint data using a deep autoencoder network. *International Journal of Geographical Information Science* 31, 1929–1951.
- Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., Zipf, A., 2020. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science* 1–27. <https://doi.org/10.1080/13658816.2020.1730848>
- Yan, Y., Feng, C.-C., Wang, Y.-C., 2016. Utilizing fuzzy set theory to assure the quality of volunteered geographic information. *GeoJournal* 1–16. <https://doi.org/10.1007/s10708-016-9699-x>
- Yan, Y., Schultz, M., Zipf, A., 2019. An exploratory analysis of usability of Flickr tags for land use/land cover attribution. *Geo-spatial Information Science* 22, 12–22. <https://doi.org/10.1080/10095020.2018.1560044>
- Yang, A., Fan, H., Jing, N., 2016. Amateur or Professional: Assessing the Expertise of Major Contributors in OpenStreetMap Based on Contributing Behaviors. *ISPRS International Journal of Geo-Information* 5, 21. <https://doi.org/10.3390/ijgi5020021>
- Zhou, Sha, Winter, S., Vasardani, M., Zhou, Shunping, 2017. Place descriptions by landmarks. *Journal of Spatial Science* 62, 47–67. <https://doi.org/10.1080/14498596.2016.1196623>
- Zielstra, D., Hochmair, H., 2011. Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Record: Journal of the Transportation Research Board* 2217, 145–152. <https://doi.org/10.3141/2217-18>
- Zielstra, D., Zipf, A., 2010. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. Presented at the In Proceedings of the 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal, pp. 1–15.

Sites Internet

Wikimapia, 2020. Wikimapia, Accessible à l'adresse : <https://wikimapia.org/>

OpenStreetMap, 2020. OpenStreetMap, Accessible à l'adresse : www.openstreetmap.org

CampToCamp, 2020, Accessible à l'adresse : <https://www.camptocamp.org/>

Visorando, 2020, Accessible à l'adresse : <https://www.visorando.com/>

Strava 2020, Accessible à l'adresse : www.strava.com

Annexe 1 : Liste de publications

Articles dans revue à comité de lecture répertoriée dans les BDI JCR, Scopus, ERIH, HCRES1

1. **Olteanu-Raimond, A.-M.**, See, L., Schultz, M., Foody, G., Riffler, M., Gasber, T., Jolivet, L., le Bris, A., Meneroux, Y., Liu, L., Poupée, M., Gombert, M., (2020). Use of Automated Change Detection and VGI Sources for Identifying and Validating Urban Land Use Change. *Remote Sensing* 12, 1186.
2. Van Damme, M.-D., **Olteanu-Raimond, A.-M.**, Méneroux, Y., (2019). Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas. *International Journal of Cartography* 5, 195–213.
3. Ivanovic, S., **Olteanu-Raimond, A.-M.**, Mustière, S., Devogele, T., (2019). A Filtering-Based Approach for Improving Crowdsourced GNSS Traces in a Data Update Context. *ISPRS International Journal of Geo-Information* 8, 380. <https://doi.org/10.3390/ijgi8090380>
4. **Olteanu-Raimond, A.-M.**, L. Jolivet, M.-D. Van Damme, T. Royer, F. Ludovic, L. See, T. Sturn, M. Karner, I. Moorthy and S. Fritz (2018) An Experimental Framework for Integrating Citizen and Community Science into Land Cover, Land Use, and Land Change Detection Processes in a National Mapping Agency, *Land*, vol. 7, n. 3, doi:doi:10.3390/land7030103.
5. Bonnel, P., E. Hombourger, **A.-M. Olteanu-Raimond** and Z. Smoreda (2017) Apports et limites des données passives de la téléphonie mobile pour la construction de matrices origine-destination, *Revue d'économie régionale et urbaine*, n. 4, pp. 647--671, doi:<http://www.revues.armand-colin.com/eco-sc-politique/revue-deconomie-regionale-urbaine/revue-deconomie-regionale-urbaine-ndeg-42017/apports-limites-donnees-passives-telephonie-mobile>.
6. Touya, G., V. Antoniou, **A.-M. Olteanu-Raimond** and M.-D. Van Damme (2017) Assessing Crowdsourced POI Quality: Combining Methods Based on Reference Data, History, and Spatial Relations, *ISPRS International Journal of Geo-Information*, vol. 6, n. 3, pp. 80, doi:10.3390/ijgi6030080.
7. Mooney, P., Minghini, M., Laakso, M., Antoniou, V., **A.-M. Olteanu-Raimond**, Skopeliti A. (2016) Towards a Protocol for the Collection of VGI Vector Data, *ISPRS International Journal of Geo-Information*, vol. 5, doi:doi:10.3390/ijgi5110217.
8. Comber, A., C. Fonte, G. Foody, S. Fritz, P. Harris, **A.-M. Olteanu-Raimond** and L. See (2016) Geographically weighted evidence combination approaches for combining discordant and inconsistent volunteered geographical information, *GeoInformatica*, vol. 20, n. 3, pp 503–527, doi:10.1007/s10707-016-0248-z.
9. **Olteanu-Raimond, A.-M.**, G. Hart, G. Touya, T. Kellenberger, G. Foody and D. Demetriou (2016) The scale of VGI in map production: A perspective of European National Mapping Agencies, *Transactions in GIS*, doi:10.1111/tgis.12189.
10. See, L., P. Mooney, G. Foody, L. Bastin, A. Comber, J. Estima, S. Fritz, N. Kerle, B. Jiang, M. Laakso, H.-Y. Liu, G. Milcinski, M. Niksic, M. Painho, A. Podor, **A.-M. Olteanu-Raimond** and M. Rutzinger (2016) Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information, *ISPRS International Journal of Geo-Information*, vol. 5, n. 5, pp. 55, doi:10.3390/ijgi5050055.
11. Bonnel, P., E. Hombourger, **A.-M. Olteanu-Raimond** and Z. Smoreda (2015) Passive Mobile Phone Dataset to Construct Origin-destination Matrix: Potentials and Limitations, *Transportation Research Procedia*, vol. 11, pp. 381--398, doi:doi:10.1016/j.trpro.2015.12.032.
12. **Olteanu-Raimond, A.-M.**, S. Mustière and A. Ruas (2015) Knowledge formalization for vector data matching using Belief Theory, *Journal of Spatial Information Science (JoSIS)*, n. 10, pp. 21-46, doi:10.5311/JOSIS.2015.10.194.

Article dans revue à comité de lecture non répertoriée dans les BDI JCR, Scopus, ERIH, AEREL

1. Bunel, M., **Olteanu-Raimond, A.-M.**, Duchêne, C., 2019b. Objets et relations spatiales composites et prise en compte du vague pour interpréter un référencement spatial indirect. *Revue Internationale de Géomatique* 29, 81–106. <https://doi.org/10.3166/rig.2019.00075>
2. **Olteanu-Raimond, A.-M.** and A. Ruas (2015) Carte de Kohonen et classification ascendante hiérarchique pour l'analyse de données géohistoriques, *Revue Internationale de Géomatique*, vol. 25, n. 4, pp. 581–599, doi:10.3166/RIG.25.581-599.
3. **Olteanu-Raimond, A.-M.**, T. Couronné, J. Fen-Chong and Z. Smoreda (2012) Modélisation des trajectoires spatio-temporelles issues des traces numériques de téléphones portables. Le Paris des visiteurs, qu'en disent les téléphones mobiles?, *Revue Internationale de Géomatique*, vol. 22, n. 3, pp. 413–437, doi:10.3166/rig.22.413-437.

Article dans revue sans comité de lecture

1. Bahoken, F. and **A.-M. Olteanu-Raimond** (2013) La construction d'une matrice de flux à partir de traces de téléphones portables, *Cartes & Géomatique - Revue du Comité Français de Cartographie*, n. 217, pp. 73-86.
2. Ruas, A., C. Plumejeaud, L. Nahassia, E. Grosso, **A.-M. Olteanu-Raimond**, M.-C. Vouloir and C. Motte (2013) GÉOPEUPLE : création et analyse de données topographiques et démographiques sur 200 ans , *Cartes & Géomatique - Revue du Comité Français de Cartographie*, n. 218, pp. 17-29.

Directions d'ouvrages

1. Foody, G. See, L., Fritz, S., Mooney, P., **A.-M. Olteanu-Raimond**, C. Fonte, V. Antoniou, (2017) *Mapping and the Citizen Sensor* book, Ubiquity Press Ltd, pp. 400, Ubiquity Press Ltd, doi:DOI: 10.5334/bbf.
2. Mallet, C., N. Paparoditis, I. Dowman, S. Oude Elberink, **A.-M. Olteanu-Raimond**, F. Rottensteiner, M.-H. Yang, S. Christophe, A. Çöltekin and M. Brédif (Editors) (2015) *Proceedings of the ISPRS Geospatial Week 2015*, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
3. Mallet, C., N. Paparoditis, I. Dowman, S. Oude Elberink, **A.-M. Olteanu-Raimond**, G. Sithole, G. Rabatel, F. Rottensteiner, X. Briottet, S. Christophe, A. Çöltekin and G. Patané (Editors) (2015) *Proceedings of the ISPRS Geospatial Week 2015*, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Ouvrages scientifiques ou chapitres

1. Follin, J.-M., J.-F. Girres, **A.-M. Olteanu-Raimond**, and Sheeren D. (2019). 'Origine de l'imperfection de l'information Géographique'. In *L'imperfection Des Données Géographiques 1*, 3:29–51.
2. Foody, G., S. Fritz, C. Fonte, B. Lucy, **A.-M. Olteanu-Raimond**, P. Mooney, L. See, V. Antoniou, H.-Y. Liu, M. Minghini and R. Vatsava (2017) Mapping and the Citizen Sensor - An overview, *Mapping and the Citizen Sensor*, pp. 1–12, Foody, G. See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. and Antoniou V. Ed., Ubiquity Press Ltd, doi:DOI: 10.5334/bbf.a
3. Mooney, P., **A.-M. Olteanu-Raimond**, G. Touya, N. Juul, S. Alvanides and N. Kerle (2017) Considerations of Privacy, Ethics and Legal Issues in Volunteered Geographic Information, *Mapping and the Citizen Sensor*, chap. Chapter 6, pp. 119-135, Foody, G. See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. and Antoniou V. Ed., Ubiquity Press, doi:10.5334/bbf.f
4. Minghini, M., V. Antoniou, C. Fonte, J. Estima, **A.-M. Olteanu-Raimond**, L. See, M. Laakso, A. Skopeliti , P. Mooney, J. Jokar Arsanjani and F. Lupia (2017) The Relevance of Protocols for VGI Collection, *Mapping and the Citizen Sensor*, vol. in press, chap. 10, pp. 223–247, Foody, G. See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. and Antoniou V. Ed., Ubiquity Press Ltd, doi:DOI: 10.5334/bbf.j
5. **Olteanu-Raimond, A.-M.** Vyron Antoniou, Cidalia Costa Fonte, Alexandra Fonseca, Magdalena Grus, Jenny Harding, Tobias Kellenberger, Mari Laakso, Marco Minghini, Andriani Skopeliti (2017) VGI in National Mapping Agencies: Experiences and Recommendations, chap. 13, *Mapping and the Citizen Sensor*, chap. Chapter 13, pp. 299-326, Ubiquity Press Ed., Ubiquity Press Ltd, DOI: 10.5334/bbf.m
6. Antoniou, V., See, L., Foody, G., Fonte, C., Mooney, P., Bastien L., Fritz, S., Liu, H.-Y., **Olteanu-Raimond, A.-M.**, M., Vatsava R., (2017), The Future of VGI, chap. 16, *Mapping and the Citizen Sensor*, DOI: 10.5334/bbf.p
7. Jolivet, L. and **A.-M. Olteanu-Raimond** (2017) Crowd and Community Sourced Data Quality Assessment, *Advances in Cartography and GIScience: Selections from the International Cartographic Conference 2017*, pp. 47–60, Peterson, Michael P. Ed., Springer International Publishing, doi:10.1007/978-3-319-57336-6_4

8. Antoniou, V., G. Touya and **A.-M. Olteanu-Raimond** (2016) Quality analysis of the Parisian OSM toponyms evolution, *European Handbook of Crowdsourced Geographic Information*, chap. 8, pp. 97--112, Cristina Capineri, Muki Haklay, Haosheng Huang, Vyrion Antoniou, Juhani Kettunen, Frank Ostermann, Ross Purves Ed., Ubiquity Press, doi:10.5334/bax.h
9. **Olteanu-Raimond, A.-M.** and T. Couronné (2014) Model the human behavior in space and time by mobile phone data, *Theories and Simulations of Complex Social Systems*, vol. 52, Dabbaghian, V., Mago, V. K. Ed., Intelligent Systems Reference Library, Springer
10. Iovan, C., **A.-M. Olteanu-Raimond, T.** Couronné and Z. Smoreda (2013) Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies, *Geographic Information Science at the Heart of Europe*, pp. 247--265, Springer
11. Smoreda, Z., **A.-M. Olteanu-Raimond** and T. Couronné (2013) Spatiotemporal data from mobile phones for personal mobility assessment, *Transport Survey Methods: Best Practice for Decision Making*, chap. 41, pp. 745--767, Transport Survey Methods: Best Practice for Decision Making, Emerald Group Publishing.

Communications avec actes dans un congrès international.

1. Bunel, M., Duchêne, C., **Olteanu-Raimond, A.-M.**, Villanova-Oliver, M., Bonhoure, G., Jouan, T., (2019). An approach for identifying and analysing reference features and spatial relations used in mountain emergency calls. Proceedings of the ICA 2, 1--8 1--8 July, Tokyo, Japon.
2. Jolivet, L., **Olteanu-Raimond, A.-M.**, Van Damme, M.-D., Gombert, M., Fauret, S. Saffroy T. (2019) Feedbacks on VGI in-situ campaign for updating LULC, Proceedings of the ICA 2, 1--8 July, Tokyo, Japon.
3. Ivanovic, S., **A.-M. Olteanu-Raimond**, S. Mustière and T. Devogele (2016) Detection of outliers in crowdsourced GPS traces , The 12th international symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Spatial Accuracy 2016 Symposium.
4. Fonte, C., L. Bastin, G. Foody, T. Kellenberger, N. Kerle, P. Mooney, **A.-M. Olteanu-Raimond** and L. See (2015) VGI Quality Control, 9th International Symposium on Spatial Data Quality (ISSDQ'2015).
5. Ivanovic, S., **A.-M. Olteanu-Raimond**, S. Mustière and T. Devogele (2015) Detection of potential updates of authoritative spatial databases by fusion of Volunteered Geographical Information from different sources, FOSS4G-Europe Conference.
6. Bonnel, P., E. Hombourger , **A.-M. Olteanu-Raimond** and Z. Smoreda (2014) Potential of passive mobile phone dataset to construct origin-destination matrix, 10th International conference on transport survey Methods: Scoping the Future while Staying on Track
7. Nabavi-Larijani, A., **A.-M. Olteanu-Raimond**, J. Perret, M. Brédif and C. Ziemlicki (2014) Investigating the mobile phone data to estimate the origin destination flow and analysis; a case study: Paris region, Proceedings of International Symposium of Transport Simulation 2014
8. Rocchini, D., **A.-M. Olteanu-Raimond**, L. Delucchi, S. Pareeth, M. Neteler and H. Nagendra (Editors) (2014) Sensing technologies and their integration with maps: mapping landscape heterogeneity by satellite imagery, Thematic Cartography for the Society, pp. 259--273, Bandrova T., Konecny M., Zlatanova S. Ed., Springer
9. Andrienko, G., N. Andrienko, G. Fuchs , **A.-M. Olteanu-Raimond** and J. Symanzik (2013) Extracting Semantics of Individual Places from Movement Data by analyzing Temporal Patterns of visits, 21 st ACM SIGSPATIAL International Conference on Advances in GIS.
10. Bahoken, F. and **A.-M. Olteanu-Raimond** (2013) Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths; The effect of spatiotemporal filtering on flow measurement, 26th International Cartographic Conference (ICC'13), International conference of ICA
11. Ruas, A., C. Plumejeaud, L. Nahassia, E. Grosso, **A.-M. Olteanu-Raimond**, B. Costes, M.-C. Vouloir and C. Motte (2013) GéoPeople: The creation and the analysis of topographic and demographic data over 200 years, Advances in Cartography and GIScience, LNG&C, Selection from 26th International Cartographic Conference (ICC'13), Buchroithner M., Prechtel N., Burghardt D. Ed., Lecture Notes in Geoinformation and Cartography.

Communications avec actes dans un congrès national.

1. Bunel, M., **Olteanu-Raimond, A.-M.** and Duchêne, C. (2018) Référencement spatial indirect : modélisation à base de relations et d'objets spatiaux vagues, *Colloque International de Géomatique et d'Analyse Spatiale (SAGEO'18)*.

2. **Olteanu-Raimond, A.-M.**, P.-A. Davoine, M. Gaio, E. Gouarderes, M.-D. Van Damme, M. Villanova-Olivier, M. Brasebin, C. Domingues, C. Duchêne, O. Favre, S. Mustière, F. Devin, Y. Le Nir, L. Moncla, S. Bouveret, J. Gensel and D. Ziebelin (2017) Projet CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne, *Actes de la conférence Sagéo 2017* (poster).
3. **Olteanu-Raimond, A.-M.** and A. Ruas (2014) Carte de Kohonen et classification ascendante hiérarchique pour l'analyse de données géohistoriques, Actes du Colloque International de Géomatique et d'Analyse Spatiale, SAGEO' 2014.
4. **Olteanu-Raimond, A.-M.**, F. Bahoken, T. Couronné and Z. Smoreda (2013) Proposition de matrices de flux temporelles issues de l'activité d'individus mobiles, Actes du Colloque International de Géomatique et d'Analyse Spatiale, SAGEO' 2013.

Communication orale sans actes dans un congrès international ou national

1. **Olteanu-Raimond, A.-M.**, LandSense : un observatoire citoyen pour le suivi de la couverture et de l'usage de sol, Sagéo'2017, Atelier « Crowdsourcing et information géographique volontaire », 6 Novembre 2017.
2. **Olteanu-Raimond, A.-M.**, C. Duchêne, M. Brasebin, C. Domingues, S. Mustière and M.-D. Van Damme (2017) From Relative to Absolute Location for Locating Victims in Mountain Area - A Preliminary Study, *Proceedings of the 28th International Cartographic Conference (ICC'17)*.
3. **Olteanu-Raimond, A.-M.**, Van Damme, M-D. (2016) Georeferencing imprecise historical names by a multi-criteria approach, 6th International Conference on Cartography & GIS, Albena, Bulgaria.
4. Foody, G., P. Mooney, L. See, N. Kerle, **A.-M. Olteanu-Raimond** and C. Fonte (2015) Current status and future trends in crowd-sourcing geographic information, AGI Foresight'2015.

Production de vulgarisation

1. Projet Choucas : Mieux localiser les victimes pour mieux les secourir, *IGN Magazine N°89*, 2018.
2. Un suivi collaboratif de la biodiversité, *IGN Magazine*, à paraître.
3. Olteanu-Raimond, A.-M., Marcuzzi, J., Van Damme, M.-D., Sturn, T., Gombert, M., Jolivet, L., See, L., Royer, T., 2020a. Crowd and community sourcing to update authoritative LULC data in urban areas. <https://doi.org/10.5281/zenodo.3691827>; jeu de données géographiques;

Annexe 2 : CV

DIPLÔMES ET DÉROULEMENT DE CARRIÈRE

Diplômes et titres universitaires

- 2005-2008 **Doctorat**, Université Paris-Est, thèse soutenue le 24 octobre **2008**
- 2004-2005 **Diplôme Universitaire** SIGmage (Systèmes d'Information Géographique et Méthodes de l'Analyse Géographique), Université de Poitiers, Département de Géographie
- 2003-2004 **Diplôme d'Etudes Approfondies** T3IA (Traitement de l'Information : Informatique, Images et Automatique), Université de Poitiers –ENSMA
- 1998-2003 **Diplôme d'Ingénieur en Télécommunications**, Ecole Polytechnique IASI, Roumanie

Déroulement de carrière

- Depuis janvier 2013 : Chargée de recherche de classe 1, Laboratoire LaSTIG, Equipe MEIG, IGN
- 2010- 2012 : Post-Doctorat, Laboratoire Sociology and Economics of Networks, Orange Labs
- 2009-2010 : Ingénieur de recherche, Procton Labs - laboratoire de recherche privé
- 2009 : Post-Doctorat, Laboratoire Image, Ville, Environnement, Université Louis Pasteur de Strasbourg
- 2005-2008 : Doctorat, Laboratoire LASTIG, Equipe COGIT, IGN.

ENSEIGNEMENT ET FORMATION À LA RECHERCHE

Enseignement

- 2019/2020 : Information géographique et qualité des données, Mastère Spécialisé®, Décision et Systèmes d'Information Géolocalisée (ENSG) (7,5h eq. TD)
- 2019/2020 : Analyse spatiale statistique, Mobilités spatiales, Intégration de données hétérogènes, qualité de données géographiques, M2IGAST, Université Paris-Est Marne la Vallée (UPEM) – Ecole Nationale des Sciences Géographiques (ENSG) (36h eq. TD)
- 2018/2019 : Information géographique et qualité des données, Mastère Spécialisé®, Décision et Systèmes d'Information Géolocalisée (ENSG) (4,5h eq. TD)
- 2018/2019 : Analyse spatiale statistique, Mobilités spatiales, M2IGAST, Université Paris-Est Marne la Vallée (UPEM) – Ecole Nationale des Sciences Géographiques (ENSG) (30h eq. TD)
- 2017/2018 : Analyse spatiale statistique, Mobilités spatiales, M2IGAST, Université Paris-Est Marne la Vallée (UPEM) – Ecole Nationale des Sciences Géographiques (ENSG) (30h eq. TD)
- 2016/2017 : Analyse spatiale statistique, Mobilités spatiales, M2IGAST, Université Paris-Est Marne la Vallée (UPEM) – Ecole Nationale des Sciences Géographiques (ENSG) (30h eq. TD)
- 2015/2016 : 4, Mobilités spatiales, M2IGAST, UPEM – ENSG (30h eq. TD)
- 2015/2016 : Analyse spatiale et qualité des données géographiques, M2IG, UPEM – ENSG (9h eq. TD)
- 2015/2016 : Module SIG, M1 Géoenvironnement, UPEM (25,5h eq. TD)
- 2014/2015 : Module SIG, M2, Géorisque, UPEM (25,5h eq. TD)
- 2014/2015 : Module SIG, M1 Géoenvironnement, UPEM (25,5h eq. TD)
- 2013/2014 : Module SIG, M2, Géorisque, UPEM (25,5h eq. TD)

Encadrement

Direction de thèse (avec dérogation HDR)

- Ibrahim MAIDANEH ABDI, thèse en cours depuis 2018 (encadrement 50%) sur le sujet : Evaluation de la qualité des données géographiques d'OpenStreetMap (OSM) par apprentissage : cas de la République de Djibouti. Thèse encadrée par Arnaud Le Guilcher (IGN/ LASTIG).

Encadrement de thèse

- Mattia Bunel, thèse en cours depuis 2017 sur le sujet : Modélisation et raisonnement spatial flous pour l'aide à la localisation de victimes en montagne. Thèse dirigée par Cécile Duchêne (IGN/ LASTIG).
- Stefan Ivanovic, thèse soutenue le 19 janvier 2018 sur le sujet : Mise à jour de référentiel géographique par fusion multi-sources de données collaboratives. Thèse co-dirigée par Sébastien Mustière (IGN/ LASTIG) et Thomas Devogele (Université de Tours).

Encadrement de post-doc

- 2019-2020 : **Lanfa Liu** ; Sujet : « Fusion d'informations hétérogènes pour la définition d'un processus de mise à jour de données de couverture et d'usage du sol », Co-encadrement avec Laurence Jolivet et Arnaud le Bris (LaSTIG)
- 2013-2014 : **Anahid Nabavi-Larijani** ; Sujet : « Modélisation et construction de flux origine-destination de piétons à partir de traces des téléphones portables ». Co-encadrement avec Julien Perret et Matthieu Brédif (LaSTIG).

Encadrement de stage

- Janv.-Fév 2018 : Stage de R&D : « Mise en place d'un outil de suivi de campagnes de collecte de données géographiques collaboratives basé sur Sémantique MédiaWiki », par Gabriel Noiret (ENSG, Paris 7, Paris 1), co-encadré avec Marie-Dominique Van Damme et Laurence Jolivet.
- Oct. 2017-Jan. 2018 : Stage d'initiation à la recherche : « Estimation de l'aire de réception des antennes téléphoniques », Doha Widadi (ENSG), Erwan Viegas (ENSG), - co-encadré avec Sébastien Mustière et Marie-Dominique Van Damme.
- Oct. 2016-Jan. 2017 : Stage d'initiation à la recherche : « Détermination automatique de l'extension spatiale d'un élément de relief en zone de montagne », par Régis Bellot (ENSG) et Maximilien Jaffres (ENSG), co-encadré avec Cécile Duchêne.
- Oct. 2016-Jan. 2017 : Stage d'initiation à la recherche : « Evaluation de sources de données de points de repère pour le repérage de victimes en montagne », par Bouthayna El-Bakkouchi et Anthony Magnetti (ENSG), co-encadré avec Cécile Duchêne.
- Fév. - Juillet 2016 : Stage de fin d'études : « Localisation d'entités nommées historiques par analyse multi-critères et prise en compte des imprécisions », par Sébastien Noguière (Université Cergy Pontoise), co-encadré avec Marie-Dominique Van Damme.
- Janv. - Fév. 2016 : Stage de recherche et développement : « Qualification des données issues de l'outil de remontées collaboratives de l'IGN », par Maryam Rhezali (ENSG, Paris 7, Paris 1), co-encadré avec Laurence Jolivet.
- Avr. - Juillet 2014 : Stage de fin d'études intitulé « Production de vues 3D pour l'aide aux secours de randonneurs », par Julian Mauborgne (AFPA), co-encadré avec Cécile Duchêne et Michael Brassebin.
- Oct. 2013-Jan. 2014 : Stage d'initiation à la recherche : « Implémentation des méthodes de l'analyse de densité de points », par Maryam Rhezali (ENSG) et Fabry Simon (ENSG), co-encadré avec Elodie Buard.
- Oct. 2013-Jan. 2014 : Stage d'initiation à la recherche : « État de l'art sur la gamification dans la saisie de données géographiques collaboratives », par Adélaïde Lorieux (ENSG) et Quy-Thy Truong (ENSG), co-encadré avec Sébastien Mustière.

- Janv. - Fév. 2014 : Stage de recherche et développement intitulé « Géolocalisation des toponymes historiques sur le référentiel géographique », par Hamza Bounouidar (ENSG, Paris 7, Paris 1).

Participation à des jurys et à des instances ou comités en lien avec l'enseignement.

- Depuis 2015 : Responsable du Module « Analyse statistiques spatiales » du Master M2IGAST, ENSG.
- 2014, 2015, Jury de soutenance de stages du Master 2 Géorisque, Université Paris-Est Marne la Vallée.

ACTIVITÉS D'ADMINISTRATION ET D'ANIMATION DE LA RECHERCHE :

- mi 2014 - mi 2016 : Co-Responsable scientifique et administratif de l'axe de recherche « Qualification et interopérabilité des référentiels de données géographiques » de l'équipe COGIT de l'IGN/LASTIG. Pendant cette période cet axe était composé de dix chercheurs (7 permanents et 3 doctorants).

Coordination groupes de recherches et de Projets.

- Depuis 2017 : Coordinatrice du projet ANR CHOUCAS impliquant les partenaires : le laboratoire LASTIG de l'IGN (porteur), le Laboratoire d'Informatique de Grenoble (LIG), le Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour (LIUPPA), ainsi que le Peloton de Gendarmerie de Haute Montagne (PGHM) de Grenoble.
- Depuis 2016 : Coordinatrice scientifique pour l'IGN dans le cadre du projet H2020, LandSense. Mon rôle dans le projet a été en premier lieu de participer au montage de projet en 2015 pendant les deux phases de soumission. Pendant la durée du projet, mon rôle a été d'assurer la coordination entre l'IGN et les partenaires du projet ainsi que le suivi financier et administratif vis-à-vis de la CE. J'ai également coordonné les travaux de recherche et d'implémentations menés par l'IGN dans le cadre du projet.
- 2016 : Collaborations avec des chercheurs de Kadaster (Pays-Bas) et NLS (Finlande). Ces collaborations, réalisées à travers différents séjours, ont permis un partage d'expertise sur les projets de recherche et innovation menés par les trois homologues autour des problématiques de la collecte collaborative et de l'éducation. Pour l'IGN, ce partage d'expertise a donné lieu à une collaboration entre l'IGN et le MNHN (Muséum National d'Histoire Naturelle) pour un travail commun autour de la plateforme Edugéo (proposé par l'IGN) et les observatoires citoyens (proposés par le MNHN).
- 2013 - 2016 : Responsable du groupe de travail WG3 « Map Production » de l'Action COST « Mapping and the Citizen Sensor ». Mon rôle a été d'animer le groupe WG3 composé de 49 membres (chercheurs et représentants de plus de 15 Agences de cartographie, homologues de l'IGN). L'animation consistait à organiser des réunions et des workshops au sein du groupe, à initier, participer et coordonner la rédaction des articles de revue et conférences impliquant des membres du groupe. J'ai été également responsable du bilan (à mi-parcours et final) de l'activité du groupe.

Participation à des projets

- Participation au montage de projet H2020 Marie Skłodowska-Curie Innovative Training Networks, **Top-Land-IQ** (Training Of the next PhD generation in cutting-edge LAND cover and land use Data management, Integration, Quality-assurance, and application). L'objectif du projet est de développer à travers la recherche l'expertise autour des données d'occupation du sol qui implique plusieurs sources de données disponibles pour produire, gérer et qualifier les données d'occupation du sol en s'appuyant sur l'intégration de données multi-sources et multi-temporelles. Dans le cadre de ce projet, j'ai la co-responsabilité d'un workpackage et le co-encadrement de trois thèses dont deux au sein de l'IGN. Le projet, soumis en janvier 2020, n'a pas été financé en 2020, malgré la bonne évaluation. Les remarques encourageantes nous ont déterminés à le resoumettre en 2021.
- Projet ANR iSpace&Time : voir Annexe 3
- Projet ANR Géopeuple : voir Annexe 3

RAYONNEMENT SCIENTIFIQUE ET DIFFUSION DE LA CULTURE SCIENTIFIQUE

Participation à des réseaux de recherche

- **Organisation de conférences et workshops :**
 - 2018 : Organisation d'une journée doctorants dans le cadre de l'Action Prospective Mobilités du GDR Magis, 24 Septembre 2018
 - 2016 : Organisation du workshop « Education and VGI », 8-9 Septembre, IGN, Paris
 - 2015 : Chair de ISSDQ' 2015 - The 9th International Symposium on Spatial Data Quality, 29-30 Septembre, La Grande Motte, France.
 - 2015 : Organisation du Management Committee Meeting pour l'Action COST TD 1202 « Mapping and Citizen Sensor », La Grande Motte, 28 Septembre, France.
 - 2015 : Co-Organisation du Workshop « Obstacles and problems for National and Regional Mapping Agencies with VGI collection activities », 26-26 Février, Bern, Suisse.
 - 2014 : Co-Organisation du Workshop « Volunteered Geographic Information – A Survey of National and Regional Mapping Agencies », 25 Mars, IGN, Paris, France.
- **ISPRS** (International Society of Photogrammetry and Remote Sensing)
 - 2016-2020 : Secrétaire scientifique de la commission WG IV/3 (Spatial data fusion and uncertainty modeling)
- **Actions COST**
 - 2013 - 2016 : Membre du Management Committee de l'Action COST TD 1202 « Mapping and Citizen Sensor »
 - 2013 – 2016 : Membre du Working Group 2 (Spatial data Quality and Infrastructures) de l'Action COST IC203 « Energic »
- **GDR MAGIS**
 - Membre de l'Action Prospective « Incertitude Epistémique ».
 - Membre actif de l'Action Prospective « Modélisation et analyse de la mobilité individuelle ».

Participation à des comités éditoriaux et expertise d'articles

- Journal of Geographical Systems (JGS), Urban Science, Journal of Spatial Information Science (JOSIS): depuis 2017
- International Journal of Geo-Information (IJGI), Cartography and Geographic Information Science (CAGIS), Elsevier Editorial System for Computer Communications : depuis 2015
- Un chapitre de livre dans *OpenStreetMap in GIScience, Experiences, Research and Applications*, Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M. (Eds.), 2015
- Deux chapitres de livre dans *Mapping and citizen Sensors*, Foody, G. See, L., Fritz, S., Mooney, P., C. Fonte, V. Antoniou (Eds.) Ubiquity Press Ltd, in press.
- Journal of Environmental Planning and Management, depuis 2014

Participation à des comités scientifiques, d'évaluation

- Jury de thèse (examinatrice): these Flora Hayat (Hayat, 2019)
- The ACM Symposium On Applied Computing, SAC' 2018, SAC'2019
- ISPRS TCIV Symposium 2018
- FOSS4G'2018
- Evaluation projets ANR : 2015, 2016 (2 projets)
- International Symposium on Spatial Data Quality: ISSDQ'2015, 2017, 2019
- XXIII ISPRS Congres 2016, XXIV ISPRS Congres 2020.

Annexe 3 : Description des projets

Géopeuple : Analyse des évolutions des maillages et des peuplements à travers l'observation du territoire, Construction des BD Géo-Historiques et approche ontologique

Le projet ANR Géopeuple a reçu le financement de l'ANR sur la période 2010-2014.

Description : L'objectif du projet a été d'analyser l'évolution sur 200 ans de la démographie en France en fonction de l'espace topographique à l'échelle de la commune. Quatre objectifs sous-jacents ont été traités : (1) construire deux bases de données démographiques et topographiques représentant la population et les changements administratifs communaux d'une part et l'espace construit et naturel de la fin du XVIIIème siècle jusqu'à nos jours, d'autre part ; (2) géo-référencer et intégrer des données historiques provenant de sources multi-dates ; (3) détecter automatiquement les symboles dans les cartes anciennes ; (4) caractériser les communes selon leurs profils topographique et démographique et analyser leur évolution démographique et topographique.

Consortium : Le projet a été coordonné par Anne Ruas de l'IFFSTAR et regroupe 4 partenaires : l'IFFSTAR, le laboratoire LaSTIG de IGN, le Laboratoire de Démographie et d'Histoire Sociale (LaDéHiS) de l'EHESS, le Laboratoire d'Informatique de Paris 6.

Site Internet : <http://geopeuple.ign.fr/>

iSpace&Time : Le SIG 4D web de la ville Vers une convergence des technologies de cartographie, grille de capteurs, réalité immersive, d'animation et de simulation

Le projet ANR iSpace&Time porté par l'IGN a reçu un financement ANR pour la période (2011-2014).

Description : L'objectif principal était de mettre en place un portail urbain sous la forme technologique d'un SIG web 4D permettant d'intégrer différentes sources de données issues des capteurs afin de développer des outils de visualisation, animation et simulation des zones urbaines en 4 dimensions. Cette plateforme permettra de répondre à plusieurs besoins tels que l'aménagement urbain, la visualisation des zones urbaines mais également des besoins de production de l'information géographique à travers, par exemple, la constitution collaborative d'une base de données pour la mobilité réduite.

Consortium : IGN, Université de Rennes 1, Institut Télécom-Sud Paris, Université Blaise Pascal – Clermont Ferrand II, France Telecom, Institut Français des sciences et technologies des transports, de l'aménagement et des réseaux.

Action Cost Mapping and the Citizen Sensors

L'action COST a reçu un financement européen pour la période (2012-2016).

Description : L'objectif du COST a été d'étudier le rôle des citoyens dans la production d'une information géographique de qualité et le lien entre les productions volontaire et traditionnelle de l'information géographique afin de proposer des protocoles et des recommandations pour la collecte de l'IGV selon différents besoins. L'action COST a été organisée en 4 groupes de travail avec un travail transversal très important : WG1 : Collecte et gestion de l'IGV ; WG2 : Compréhension et formation des volontaires afin de leur apprendre les bonnes pratiques ; WG3 : Production traditionnelle de l'IG afin d'identifier les besoins pour la production de l'IG d'autorité ; WG4 : La validation et la qualification de l'IGV.

Consortium : Portée par l'Université de Nottingham (Gilles Foody) ; 28 pays membres

Site web: [TD1202](https://td1202.com)

LandSense : A Citizen Observatory and Innovation Marketplace for Land Use and Land Cover Monitoring

Le projet Landsense a reçu un financement pour la période 2016-2020 lors de l'appel à projet du programme Horizon 2020 de la Commission Européenne (CE).

Description. L'objectif du projet est de proposer un observatoire citoyen pour le suivi de l'occupation et l'usage du sol. Il s'agit plus particulièrement de construire un observatoire citoyen qui permettra d'accéder à différentes sources de données : données d'autorité (produites à l'échelle nationale ou européenne), données ouvertes, données produites par les citoyens, d'une part et à des outils de collecte de données, d'autre part. Un des objectifs du projet est d'impliquer davantage le citoyen dans le suivi de l'occupation du sol et de démontrer l'apport et la qualité de la remontée citoyenne. Cette implication se fera à travers trois campagnes de collecte de données réalisées en même temps dans différents pays, et s'intégrant autour de trois problématiques : suivi du changement des zones urbaines et rurales (France, Allemagne et Autriche), suivi de l'usage des terres agricoles (Serbie et Espagne) et suivi de l'habitat et des zones protégées (Espagne, Indonésie). L'IGN est impliqué dans la mise en place de la campagne de suivi de changements en zone urbaine et péri-urbaine afin de mesurer la consommation des espaces naturels et forestiers et l'intégration des changements détectés dans la base de données Occupation du Sol Grande Echelle (OCS-GE).

Consortium : Le projet est coordonné par Steffen Fritz de l'International Institute for Applied Systems Analysis (IAASA, Autriche) et regroupe 17 partenaires (Birdlife International, Sinergise, University of Nottingham, InoSens doo, GeoVille Information, Systems GmbH, Environment Agency Austria, IGN-France, European Citizen Science Association, StZ Felis, University of Heidelberg, Wageningen University, VU University Amsterdam, Joint Research Centre, Secure Dimensions, Friends of the Earth/Global 2000, City of Heidelberg)

Site web: <https://landsense.eu/>

Choucas : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne

Le projet Choucas a reçu le financement de l'ANR sur la période 2017-2021.

Description : Le projet CHOUCAS répond à un besoin exprimé par le Peloton de Gendarmerie de Haute Montagne (PGHM) de Grenoble. La localisation de la victime pendant la phase d'alerte repose sur un dialogue entre le requérant (personne qui donne l'alerte, qui peut être la victime ou un tiers) et les secouristes. Plusieurs éléments compliquent cette tâche de localisation. Le requérant, qui connaît souvent mal la montagne, décrit la position de la victime et l'itinéraire suivi à l'aide de positions relatives imprécises par rapport à des objets repères eux-mêmes souvent ambigus. Les données nécessaires aux secouristes pour la localisation sont hétérogènes, multi-sources, réparties et insuffisamment structurées pour être interrogées au mieux.

Le projet CHOUCAS a pour objectif de proposer des méthodes et des outils permettant de constituer et d'enrichir des données géographiques issues de sources hétérogènes, de définir des modèles de raisonnement spatial flou et de visualisation de données imparfaites afin de favoriser le processus de décision dans le contexte de la localisation des victimes en milieu naturel terrestre. L'enjeu est donc d'améliorer la capacité de réponse à une alerte et de sauver des vies.

Consortium : Le projet est coordonnée par Ana-Maria Olteanu-Raimond et regroupe les partenaires suivants : le laboratoire LaSTIG de l'IGN, l'équipe STEAMER du LIG de l'Université de Grenoble, le laboratoire UPPA de l'Université de Pau, et le PGHM de Grenoble.

Site web : <http://choucas.ign.fr/>