

RETOURS D'UNE CAMPAGNE IN-SITU DE VGI POUR LA MISE À JOUR DE DONNÉES D'OCCUPATION DU SOL

par Laurence Jolivet, Anna-Maria Olteanu-Raimond, Marie-Dominique Van Damme, Marie Gombert, Simon Fauret et Thierry Saffroy

Université Paris-Est, LASTIG, IGN, ENSG,
94160 Saint-Mandé
prénom.nom@ign.fr

Mots-clefs : Information géographique volontaire, données d'occupation du sol, campagne in-situ, mise à jour des données

L'article présente un travail réalisé dans le cadre du projet européen H2020 2016-2020 LandSense. Le projet a pour but de mettre en place une plateforme collaborative pour le suivi et la mise à jour de données d'occupation du sol, en information sur la couverture et sur l'usage. Cette plateforme vise à permettre la remontée d'informations en occupation du sol par les citoyens ainsi que par des organisations publiques et des entreprises privées. Elle offre l'accès à des bases de données existantes et propose un ensemble de services, par exemple un service en ligne de détection automatique de changements d'occupation du sol entre deux dates (Matheus., 2018). L'un des cas d'étude pilote dans LandSense est le suivi des dynamiques urbaines et péri-urbaines en proposant de compléter les données d'autorité existantes.

Dans ce contexte, un des objectifs de l'IGN (Institut national de l'information géographique et forestière) est d'étudier le potentiel de l'information géographique volontaire (ou VGI) (Goodchild, 2007) pour l'enrichissement et la mise à jour des données d'autorité d'occupation du sol. Cette approche fait appel à l'implication des citoyens ainsi qu'à celle d'un certain nombre d'entités d'autorité publique, notamment pour participer à la collecte d'informations et pour assembler et cartographier ces informations. La base de données ciblée est ici l'OCS GE (pour occupation du sol à grande échelle) produite par l'IGN à l'aide de bases existantes et de méthodes semi-automatiques pour une utilisation à l'échelle du 1 : 5.000ème. Une des spécifications importantes de l'OCS GE est qu'elle contient des polygones (vecteurs) adjacents, chacun renseigné par deux attributs distincts pour la couverture et pour l'usage du sol. Deux nomenclatures de couverture et d'usage ont en effet été définies. Elles comprennent respectivement

quatre et trois niveaux hiérarchiques et actuellement 14 et 17 classes de niveaux pour le plus détaillé. Cependant, certaines classes ont été fusionnées par rapport à la nomenclature théorique préalable à cause de la difficulté de pouvoir les distinguer lors de leur construction semi-automatique. C'est en particulier le cas des classes d'usage du sol, telles que la classe réunissant les usages d'activités secondaires, d'activités tertiaires et résidentiel. Par ailleurs, l'OCS GE est une base millésimée, c'est-à-dire que sa publication correspond à la réalité du terrain lors d'une année en particulier et qu'il n'y a pas de version intermédiaire entre les publications (millésimes). Sa prévision de rythme de publication est tous les trois ans. Les besoins de l'OCS GE sont d'une part la possibilité d'enrichissement de la nomenclature actuelle et d'autre part la détermination et la prise en compte des changements pour le millésime suivant. Ces besoins peuvent être synthétisés en trois points principaux :

- améliorer les données d'occupation du sol en découpant les classes qui ont été fusionnées afin de correspondre davantage à la nomenclature initiale ;
- détecter les changements en occupation du sol afin qu'ils puissent être plus rapidement intégrés à la prochaine publication, prévue pour l'année 2019 ;
- détecter les changements (pour le millésime suivant) ou les mises à niveau (pour des millésimes historiques) en occupation du sol entre les deux versions existantes de l'OCS GE en 2013 et en 2016.

Pour déterminer des réponses aux besoins de cette donnée d'autorité, une campagne VGI a été établie afin de collecter des informations par observations directes sur le terrain. La stratégie de campagne, les concepts utilisés et les outils collaboratifs sont présentés dans Olteanu-Raimond (2018). En particulier, une application pour téléphone mobile nommée PAYSAGE a été développée afin de recueillir les contributions sur le terrain et un wiki sémantique accompagné d'un forum utilisateur a été déployé dans le but de partager des informations sur les données et sur les procédures de contributions. Cet article se concentre à présent sur la mise en œuvre de la stratégie de campagne et sur son déroulement via une analyse des données collectées.

Les actions collaboratives lors de cette campagne VGI in-situ concernaient cinq tâches précises listées ci-dessous, pouvant potentiellement correspondre à l'un des trois besoins identifiés précédemment.

- 1/ L'ajout de nouvelles informations concernant les bâtiments présents dans une des bases de données utilisées pour construire les polygones de l'OCS GE. Ces informations sont l'usage principal, l'usage secondaire et le nombre d'étages. Une liste de valeurs d'usage prédéfinies est proposée aux contributeurs, identique pour les usages principal et secondaire.
- 2/ La mise à niveau des zones agricoles. Cette tâche porte sur l'amélioration de la précision sémantique des polygones correspondant à un usage agricole. Certains de ces polygones peuvent en effet correspondre à plusieurs usages et il s'agit d'indiquer si l'usage principal est bien agricole (par exemple des champs cultivés) ou non (par exemple des bâtiments ou des espaces résidentiels).
- 3/ La mise à jour de l'activité des carrières. Cette tâche se concentre sur les polygones de l'OCS GE dont la classe d'usage en 2013 est celle des activités d'extraction. Les contributeurs indiquent si la carrière est encore en activité, si elle est fermée ou alors si le lieu est abandonné. Cela doit permettre ensuite de confirmer ou de modifier la classe d'usage à la date de la contribution.
- 4/ La mise à jour des zones de construction. Les polygones concernés sont ceux de l'OCS GE 2013 dont la classe est celle des zones en transition qui correspond à des travaux en cours. Si les travaux sont terminés, les contributeurs éditent et enregistrent les valeurs des classes de couverture et d'usage.

5/ La validation des changements détectés par le service de la plateforme LandSense CDS (pour). Les sorties de ce service sont des pixels localisés ayant été reconnus comme changements potentiels entre deux images Sentinel-2 de dates différentes. Un type de changement est attribué parmi la liste suivante: industriel, infrastructure, résidentiel et autre. Les contributeurs sont amenés à confirmer ou infirmer le changement. Si un changement a bien eu lieu, les nouvelles valeurs de classes de couverture ou d'usage sont éditées puis enregistrées en base.

La campagne sur le terrain s'est déroulée dans la région Occitanie en France pendant quatre mois entre juillet et octobre 2018. Les sites ciblés étaient en particulier la ville de Toulouse, la métropole de Toulouse et le Parc Naturel Régional des Grands Causses. Dans ces sites, des localisations avaient été sélectionnées au préalable car correspondant aux thématiques des tâches définies. Les localisations sont de géométrie ponctuelle mais correspondent à l'ensemble du polygone de l'OCS GE respectif concerné (centroïde, décalé ensuite si besoin à l'intérieur du polygone). Au total, 60 321 localisations ont été importées dans les outils collaboratifs et ont donc été destinées à être visitées par des contributeurs. La mise en œuvre sur le terrain consistait pour les contributeurs à utiliser l'application mobile pour se déplacer vers les localisations puis d'éditer en ligne les informations en suivant les instructions écrites dans l'interface. Les contributeurs attendus étaient en particulier des citoyens déjà impliqués ou non dans des initiatives collaboratives, des membres des autorités publiques et des étudiants dans le champ disciplinaire de la géomatique. L'IGN s'est occupé de lancer la campagne, puis celle-ci a été encadrée à la fois par les universitaires, des autorités territoriales locales et l'IGN. La communication autour de la campagne, en amont et pendant son déroulement, a été effectuée via les comptes institutionnels sur les réseaux sociaux, par des communications mail, dans des conférences et au cours de réunions entre professionnels.

Le bilan quantitatif de la campagne est le suivant : 331 contributions correspondent à 277 localisations et ont été effectuées par 26 contributeurs (plusieurs localisations pouvant être visitées par des contributeurs différents). L'édition des informations sur les bâtiments (tâche 1) a fait l'objet de 178 contributions, les mises à jour et à niveau de l'usage du sol (tâches 2, 3 et 4) à 137 contributions et la validation des changements issus du CDS (tâche 5) à 16 contributions (tableau 1).

Thème	Nombre de contributions	Nombre de localisations	Nombre de contributeurs
Informations sur les bâtiments	178	135	21
Mise à jour et à niveau de l'usage du sol	137	128	18
Validation des changements CDS	16	14	10

Tableau 1 : Synthèse du nombre de contributions ventilées par thème correspondant aux tâches, et associé aux nombres de localisations et de contributeurs.

Nous détaillons les contributions par tâche. Nous nous sommes attachés à décrire le contenu des contributions ainsi qu'à les qualifier à partir de vérification de la cohérence intrinsèque et extrinsèque (Senaratne et al. 2016). Tout d'abord, la cartographie des contributions permet de se rendre compte de leur répartition spatiale (fig. 1). Cette répartition est caractérisée par des clusters de localisations. Ces regroupements sont principalement dus au mode de contribution sur le terrain. En effet, les contributeurs s'étant déplacé à pied en majorité ou en voiture ont eu tendance à visiter des localisations proches spatialement afin d'optimiser le rapport entre temps de trajet et nombre de contributions.

Concernant la tâche pour les informations attributaires des bâtiments, les contributeurs ont édité les valeurs à partir de liste prédéfinie. Ces valeurs étaient initialement vides car elles ne sont pas présentes dans les bases de données existantes. Il n'est donc pas possible de comparer avec des valeurs antérieures et donc d'estimer les changements survenus entre deux dates. Pour l'usage principal, 20 valeurs ont été sélectionnées sur les 57 prédéfinies, majoritairement concernant un usage résidentiel ou commercial. Pour l'usage secondaire, 103 contributions soit 58 % n'en mentionnaient pas. Lorsque l'usage secondaire est distinct de l'usage principal, la majorité ont des valeurs similaires, par exemple « bâtiment d'habitation, maison » et « immeuble, collectif » ou encore « église » et « cathédrale ». Certains usages secondaires paraissent toutefois peu similaires ou rarement associés avec les usages principaux indiqués, par exemple « usine hydroélectrique » en usage principal et « abbaye » en second. Par ailleurs, l'étude des contributions effectuées par des contributeurs différents sur une même localisation met en évidence des variations. Par exemple, un même bâtiment peut correspondre à un usage unique ou bien à deux usages distincts

selon les contributions. Pour la qualification de ces contributions, nous avons donc étudié la cohérence entre les éditions d'une contribution (entre les usages principal et secondaire), la cohérence entre les éditions par des contributeurs différents (concernant 28 localisations) et la cohérence avec des bases de données extérieures (par exemple des informations de la BD TOPO, OSM, images aériennes) (Fonte et al. 2017).

Les tâches de mises à jour et à niveau de l'usage du sol de certains polygones de l'OCS GE impliquaient la modification de valeurs initiales, d'usage mais également de couverture si concernée. Sur les 137 contributions, 79 concernent les zones de construction, 50 les zones agricoles et 8 l'activité des carrières. Il y a eu 92 % des localisations avec des changements entre valeurs initiales et contributions.

Pour la tâche de mise à niveau ou à jour de l'usage des zones agricoles, les contributions n'identifient aucun changement en couverture du sol. Toutes les valeurs d'usage du sol ont par contre été modifiées en « résidentiel » (exemple en figure 2). Qualifier ces informations n'est cependant pas aisé. Il est possible de comparer les valeurs entre les contributeurs pour les localisations à plusieurs contributions, mais pour cette tâche cela ne concerne que trois localisations. Nous remarquons également que les parcelles agricoles concernées sont entourées par des polygones d'usage résidentiel. Cela soulève la question de l'adéquation entre l'échelle spatiale d'observation et celle des spécifications. Les contributeurs peuvent également avoir une interprétation personnelle des paysages et donc de l'usage des sols.

Les localisations concernant les carrières ont été sélectionnées au préalable sur des polygones à couverture « zones à matériaux minéraux » et à usage « activités d'extraction ». Les contributions pour la tâche de mise à jour de leur activité ne correspondent à aucun changement en couverture mais tous sont modifiés en « sans usage ». La qualification de ces données pourra par exemple être envisagée en croisant les informations des contributions avec des bases de données régionales mises à jour sur les carrières en activité et leurs entreprises exploitantes.

Les zones de construction (ou zones de transition), thème correspondant à l'usage initial, ont un pourcentage de changement égal à 86 % soit en usage ($n = 65$), soit en couverture ($n = 60$) soit pour les deux ($n = 57$). La grande majorité des changements en couverture sont constatés à 57 % de « zones à matériaux minéraux » et « formations herbacées » en « zones bâties ». Les changements en usage le sont en « résidentiel », « zones de transition » et « sans usage ».

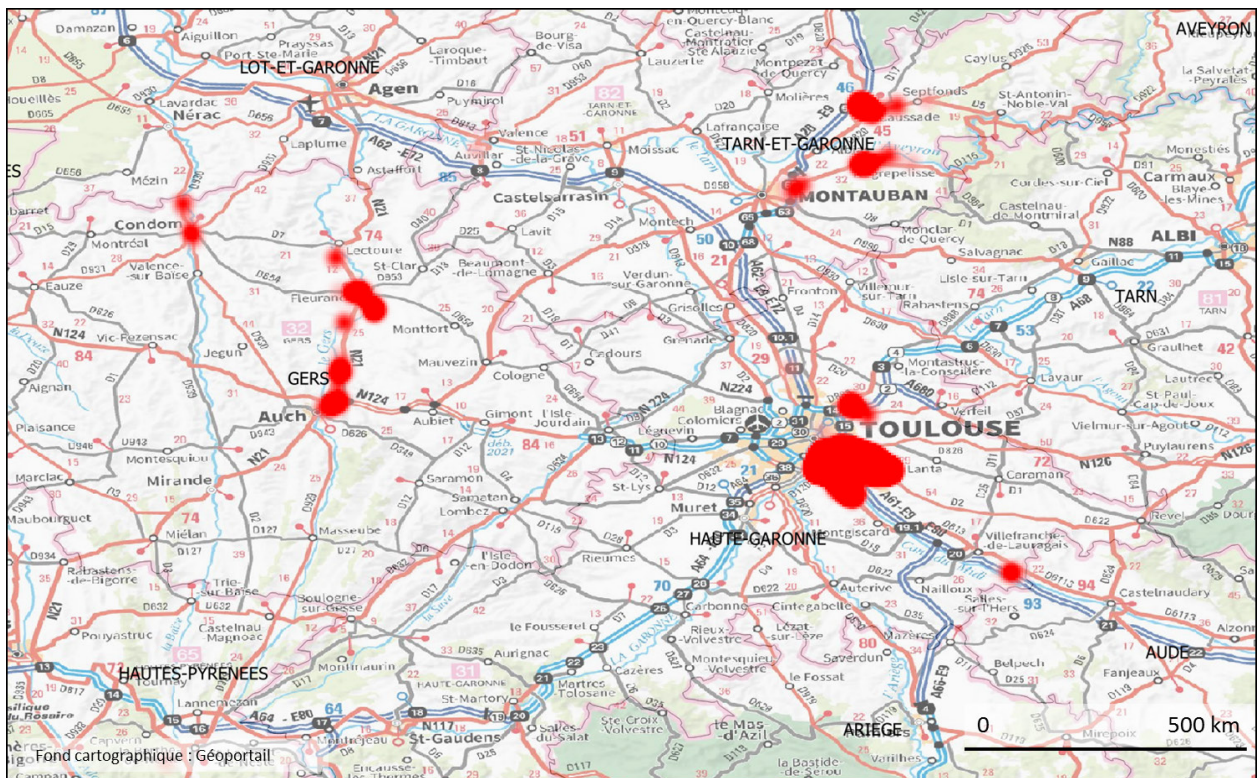


Figure 1 : Carte de chaleur des localisations des contributions : leur répartition spatiale en clusters est propre au regroupement spatial des visites sur le terrain par les contributeurs.

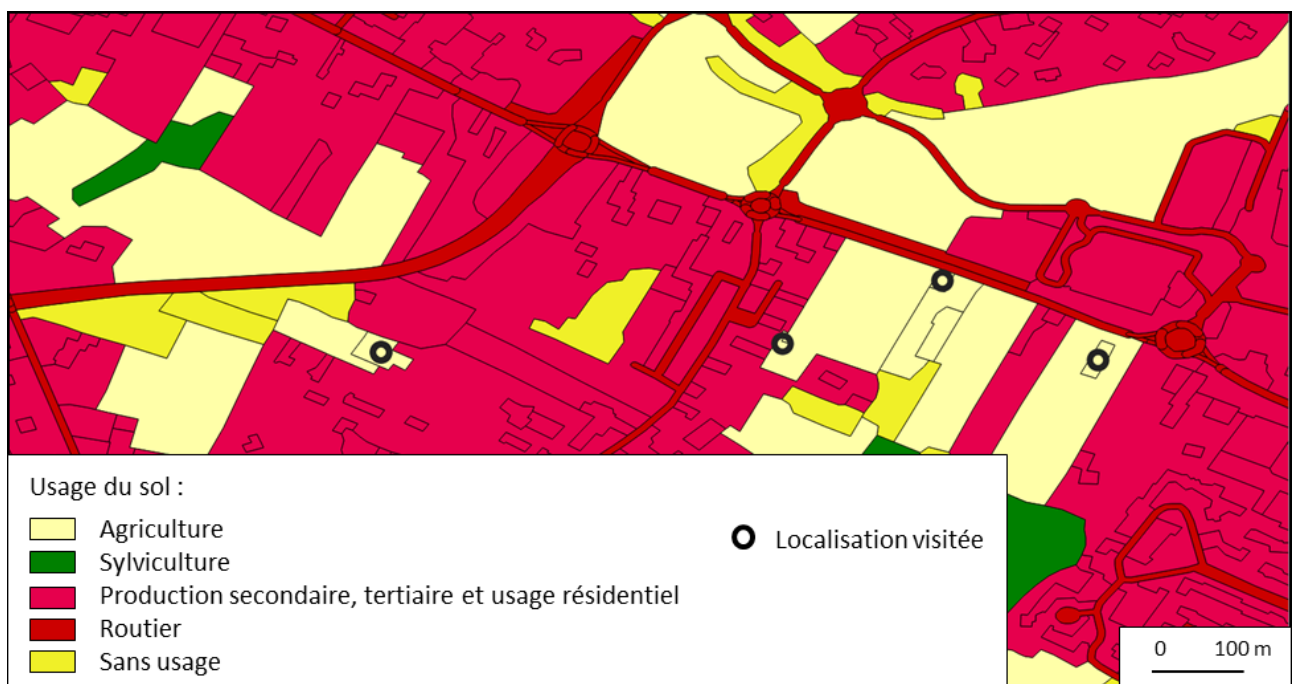


Figure 2 : Exemple de quatre localisations visitées et pour lesquelles les contributeurs ont indiqué un changement de l'usage « agriculture » à « résidentiel ».

Enfin, la tâche de validation des changements issus du CDS a impliqué une première contribution en valeurs booléennes et, en cas de changement validé, une édition du nouvel usage. Parmi les 14 localisations visitées de changements (pour 16 contributions), 7 avaient été reconnues initialement par le CDS comme de type industriel et les 7 autres comme résidentiel. Toutes les contributions valident l'existence d'un changement sur le terrain par rapport à la version de l'OCS GE visualisée (81 % à la fois en couverture et en usage). Toutes les couvertures sont à présent indiquées en « zones bâties », ce qui paraît cohérent avec les types de changements détectés par l'algorithme. Les couvertures initiales correspondantes étaient « formations herbacées », « formations arbustives et sous-arbrisseaux » et « peuplement de feuillus ». Toutefois deux contributions n'indiquent pas de changements en couverture (« zones bâties » initialement vers « zones bâties ») mais uniquement en usage par précision de celui-ci (« production secondaire, tertiaire et usage résidentiel » vers « commerces ») ou par changement (« production secondaire, tertiaire et usage résidentiel » vers « autre production primaire »), cet usage pouvant être difficile à calculer pour un algorithme.

Par ailleurs des éditions de valeurs d'occupation du sol, 434 photos ont été prises par les contributeurs sur le terrain. Elles sont associées à 322 contributions et à 262 localisations visitées. Des commentaires en texte libre ont été écrits pour 53 contributions. Ceux-ci apportent des explications concernant la valeur choisie ou validée en occupation du sol (par exemple, précision de l'usage par rapport à la nomenclature proposée, toponymie) et contiennent des indications sur la certitude estimée par le contributeur des informations données (par exemple, sur le nombre de ce qui est considéré comme étage dans un bâtiment, ou bien liées aux conditions d'observation sur le terrain comme le manque de visibilité).

Les résultats de cette première campagne in-situ du projet LandSense sont prometteurs. Ils montrent que des contributions à partir d'observations sur le terrain apportent des informations intéressantes pour l'aide à l'amélioration et à la mise à jour des

données d'autorité en occupation du sol. À venir, la recherche continuera dans un objectif de qualification systématique des contributions : l'étude des photos et des commentaires sera poursuivie et des bases de données autres seront mobilisées (Li et Goodchild, 2012). Une chaîne de traitements sera proposée afin de prendre en compte les informations relevées sur le terrain en complément d'autres sources d'informations sur la mise à niveau et mise à jour de l'occupation de sol et de les intégrer à l'OCS GE selon les spécifications de la base. Concernant les retours sur le déroulement de la campagne, ils ont été positifs sur les outils et interfaces déployés car mentionnés comme intuitives et conviviales. Toutefois, il est apparu que les informations de couverture et d'usage du sol ne sont pas évidentes à expliquer par rapport à une réalité terrain parfois complexe, et qu'il peut être difficile de relier les consignes en spécifications et leurs observations (Comber 2008). Il n'a pas toujours été aisé de choisir la valeur d'usage ou de couverture semblant adaptée. Une des conclusions les plus importantes de cette campagne est l'importance que peut avoir une communauté de citoyens prêts à participer à de tels projets collaboratifs. Lors de cette campagne, celle-ci a été limitée car il est difficile de créer une communauté autour du thème de l'occupation du sol et la participation la plus forte a été constatée pour des personnes déjà habituées à ce type de campagnes ou d'informations. Une seconde campagne dans le cadre de LandSense sera menée en 2019 et début 2020. Une attention particulière à la communication sera portée autour de la campagne et à l'animation de la communauté pendant son déroulement. Des événements pourront être organisés comme des *mapathons* (Schultz *et al.*, 2017), ce qui peut provoquer une émulation et ce qui permettrait de préciser les besoins et les attentes, par exemple concernant la nomenclature. Des défis peuvent aussi être lancés concernant des thématiques particulières d'occupation du sol ou un site géographique restreint. Ceci permettrait aux contributeurs de concentrer leurs efforts sur des tâches précises. Enfin, la communication auprès des contributeurs de l'ensemble des contributions et de leur prise en compte est un aspect important à mettre en place.

Bibliographie

Comber A. J. 2008 "The separation of land cover from land use using data primitives". *Journal of Land Use Science*, 3(4), p. 215–229, doi:10.1080/17474230802465173.

Fonte C., Patriarca J., Minghini M., et al., 2017, "Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application." Dans Calazans Campelo, C. E., et al. (Eds), *Volunteered Geographic Information and the Future of Geospatial Data..* Hershey, PA, USA, IGI Global, (Advances in Geospatial Technologies)

Goodchild M. F., 2007, "Citizens as sensors: The world of volunteered geography". *GeoJournal*, 69, p. 211–221.

Li L., Goodchild M. F. 2012, "Constructing places from spatial footprints". *GEOCROWD '12*, Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, p. 15–21.

Matheus A., Albrecht F., Batic M., Perger C. 2018, Deliverable 3.3: LandSense Engagement Platform I. Available online at: <https://lep.landsense.eu/Project/Deliverables>.

Olteanu-Raimond A.-M., Jolivet L., Van Damme M.-D., et al.. 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>.

Schultz M., Voss J., Auer M., et al. 2017, "Open land cover from OpenStreetMap and remote sensing". *International Journal of Applied Earth Observation and Geoinformation*, 63, p. 206–213

Senaratne H., Mobasheri A., Loai Ali A., et al. 2016 "A review of volunteered geographic information quality assessment methods", *International Journal of Geographical Information Science*, doi:10.1080/13658816.2016.1189556.