

## Offre de thèse - 2022



### Mise à jour de données d'occupation du sol : une approche multi-source et multi-modale pour la qualification de changements et la caractérisation des usages du sol par apprentissage profond



IGN – LaSTIG, 73 avenue de Paris 94160 SAINT-MANDÉ

#### Objet de la thèse

Cette thèse vise à proposer des approches de qualification et de mise à jour des données d'occupation du sol, particulièrement concernant l'usage, en s'appuyant sur l'intégration de données hétérogènes multi-sources et multi-modales telles que des données de télédétection, des bases de données vecteur provenant d'institutions, de la recherche ou d'études et produites à un niveau national ou local, ainsi que l'information géographique volontaire ou involontaire issues d'initiatives collaboratives.

#### Descriptif de la thèse

La connaissance de l'occupation du sol (OCS), incluant sa couverture (description physique de l'espace) et son usage (son utilisation anthropique), à une échelle spatiale et temporelle fine et avec une sémantique détaillée, permet de suivre les évolutions du territoire. Des données d'occupation du sol à grande échelle spatio-temporelle sont en effet nécessaires au suivi de phénomènes tels que l'artificialisation des sols, un enjeu majeur pour la France et l'Europe car constituant une des premières causes du changement climatique et de l'érosion de la biodiversité. À partir de ces données, des états des lieux annuels de la consommation d'espaces (Action 7 du Plan biodiversité 2018<sup>1</sup>) peuvent être publiés et des indicateurs calculés (ex. *SDG Indicators*<sup>2</sup>). Pour répondre à ce besoin, des acteurs publics dont l'IGN produisent des bases de données millésimées d'occupation de sol : CORINE Land Cover (CLC) et Urban Atlas au niveau européen, OCcupation du Sol à Grande Échelle (OCS GE<sup>3</sup>) en France. Ces produits présentent néanmoins des limites quant à leur actualité (ex. millésime disponible tous les 6 ans pour CLC, tous les 3 ans pour OCS GE) et à leur précision sémantique (ex. combinaison entre couverture et usage dans CLC, une même classe pour les trois groupes d'usages résidentiel, industriel et tertiaire dans OCS GE). L'analyse automatique de données de télédétection (imagerie aérienne et satellite) permet d'accéder à la couverture du sol. Au fur et à mesure de la constitution initiale des bases de données de couverture, des jeux d'exemples de plus en plus conséquents sont constitués et sont à même d'améliorer l'automatisation de la production et de la mise à jour en utilisant des méthodes d'apprentissage profond, comme c'est le cas pour l'OCS GE. En revanche, l'usage du sol reste une information plus difficile à extraire des données de télédétection. La mise en place d'outils permettant d'intégrer différentes sources d'observations faciliteraient *a priori* la cartographie du territoire, sa caractérisation et sa mise à jour.

<sup>1</sup> <https://www.ecologie.gouv.fr/plan-biodiversite> ; <https://biodiversite.gouv.fr/>

<sup>2</sup> <https://unstats.un.org/sdgs/indicators/indicators-list/>

<sup>3</sup> <https://geoservices.ign.fr/ocsge>

L'objectif général du projet de thèse est de proposer une approche permettant de **produire des données d'occupation du sol multi-temporelles, fiables, et respectant le principe FAIR** (*Findable, Accessible, Interoperable, Reusable*)<sup>4</sup>. Trois objectifs scientifiques en sont déclinés :

1. analyser, sémantiser et qualifier des changements géométriques et sémantiques par une approche multi-source et multi-modale ;
2. intégrer ces changements pour mettre à jour des données d'usage du sol existantes et produire des millésimes annuels ;
3. qualifier les millésimes produits.

Ce sujet s'intéresse ainsi principalement à la mise à jour de données existantes grâce à l'apport de données multi-sources et multi-modales. Les données visées sont similaires à l'OCS GE de l'IGN, relativement à ses spécifications définies, remplies ou non (typologie des classes de couverture et d'usage, superficies minimales de prise en compte des objets). Les données mobilisées seront des données aériennes produites par l'IGN (BD ORTHO, modèles numériques de surface, acquisition LiDAR), des images satellitaires (Pléiades, SPOT, Sentinel-2), des bases de données vecteur institutionnelles (dont la BD TOPO), des données produites à l'échelle locale (ex. observations collectées dans et par les parcs naturels), des données issues d'initiatives collaboratives et correspondant à de l'information géographique volontaire (ex. OpenStreetMap) ou involontaire (ex. Foursquare).

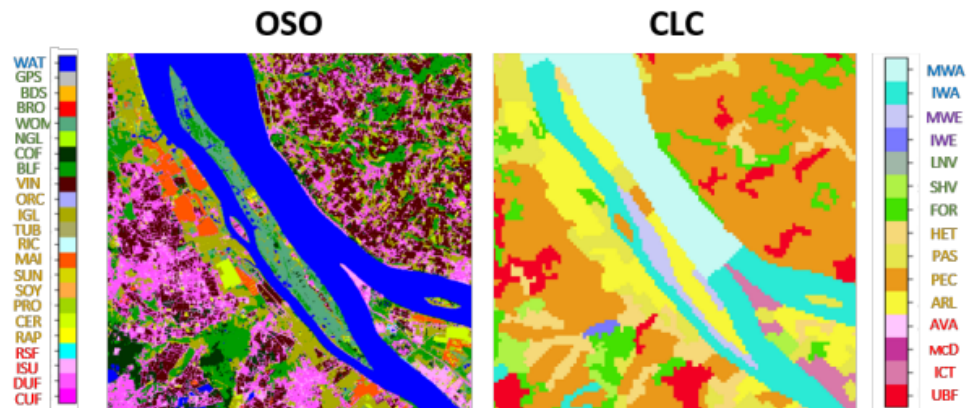
L'approche envisagée s'appuyera sur des méthodes d'apprentissage, notamment d'apprentissage profond, à même d'intégrer de manière implicite les spécifications des données souhaitées (c'est-à-dire en les apprenant à partir d'exemples plutôt qu'en les formalisant sous forme de règles d'experts). Des méthodes d'analyse spatiale plus traditionnelles seront également utilisées de manière complémentaire ; leurs résultats pourront constituer des données intermédiaires en entrée des méthodes d'apprentissage.

L'apprentissage profond a déjà été utilisé avec succès ces dernières années avec la mise en oeuvre de plusieurs architectures de réseaux de neurones convolutifs pour la sémantisation de l'occupation du sol (Gaetano et al., 2018 ; Carvalho et al., 2019 ; Stocker et al., 2020) et la détection de changements (Papadomanolaki et al., 2019 ; Peng et al., 2019 ; Pomente et al., 2018 ; Lebedev et al., 2018) à partir de données de télédétection. Dans le cas de la détection de changements, l'entraînement des algorithmes s'effectue à partir de changements avérés et annotés. Si ces algorithmes sont adaptés au traitement de données en format raster, l'exploitation directe de données en format vecteur par apprentissage profond comme une alternative à l'analyse spatiale traditionnelle est restée peu développée. Par ailleurs, les résultats fournis par ces méthodes automatiques nécessitent encore, malgré leur bonne qualité, un effort important de qualification et de correction manuelle. La caractérisation automatique des types d'usage du sol à partir de données de télédétection reste également difficile. Toutefois, des travaux ont été menés sur un sujet proche concernant la dérivation d'une carte d'OCS (correspondant aux spécifications de CLC dont la légende mêle parfois couverture et usage) à partir d'une carte existante de légende différente et d'échelle spatiale plus petite (carte OSO – Occupation des SOIs – produite par Theia), ceci à l'aide d'une architecture profonde de type U-NET (Baudoux et al., 2021).

---

<sup>4</sup> <https://www.nature.com/articles/sdata201618>

Des travaux récents basés sur la modélisation de connaissances d'experts et la validation participative (Olteanu-Raimond et al., 2020 ; Liu et al., 2020) ou l'enrichissement sémantique (Fonte et al., 2017) ont montré que l'intégration de l'information géographique volontaire ou involontaire permet de mettre à jour et d'améliorer les données de couverture des sols. Cependant, la fusion par apprentissage profond entre données de télédétection (brutes ou déjà sémantisées) et données issues de bases vecteur (bases de données topographiques, signalements participatifs) pour la dérivation de l'usage, a été peu étudiée.



*Exemple de génération automatique d'une carte d'occupation du sol généralisée de type CLC à partir d'une carte de couverture du sol (produit Théia OSO) [Baudoux et al., 2021]*

**L'originalité du projet de thèse tient à l'approche globale et à l'utilisation d'informations géographiques volontaires et involontaires.** L'approche envisagée pour mettre à jour des données d'OCS s'éloigne en effet de la méthode de production traditionnelle des produits d'OCS publiés. Elle s'attachera à associer des sources en télédétection à des sources en bases vecteur structurées et non-structurées, et à exploiter ces différentes sources par des processus combinant apprentissage et analyse spatiale. L'intégration de sources d'information géographique volontaire et involontaire est un défi à relever dû à la difficulté de qualification des données, mais prometteur pour améliorer la caractérisation de l'usage du sol.

Les avancées attendues sont :

1. des méthodes de qualification des changements selon la précision géométrique et sémantique, et leur pertinence pour la mise à jour de l'occupation du sol ;
2. une méthode permettant d'améliorer la description de l'usage du sol en mobilisant l'information géographique volontaire et involontaire en plus des données institutionnelles ;
3. une méthode de mise à jour des classes d'occupation du sol prévoyant à la fois la modification de la géométrie et celle du type d'usage ;
4. une démarche pour constituer un jeu de données d'entraînement pour les méthodes d'apprentissage ainsi qu'un jeu de données de validation, qui s'inspirerait des mécanismes collaboratifs.

La thèse proposée impliquera donc de développer des méthodes automatiques à même de croiser différentes sources d'information disponibles, ceci pour identifier les changements potentiels du territoire, caractériser leur nature (géométrique et sémantique), qualifier leur pertinence et proposer leur intégration dans une base de données d'occupation du sol. Les questions scientifiques traitées relèveront donc de la fusion de données et de l'analyse

spatiale, qui seront ici abordées en combinant des approches traditionnelles de traitement des données en format raster (classification de données de télédétection, extraction d'information) et vecteur (analyse spatiale, qualité et intégration) ainsi que des techniques innovantes fondées sur des méthodes d'apprentissage profond.

### Modalités pratiques

- Les travaux se dérouleront dans les locaux du LASTIG à l'IGN à Saint-Mandé (94). Le contrat doctoral, de trois ans, pourra être réalisé avec ou sans charge de cours, selon le profil et le souhait du candidat et les besoins de l'Ecole Nationale des Sciences Géographiques. La thèse sera dirigée par Ana-Maria Olteanu-Raimond, et co-encadrée par Laurence Jolivet et Arnaud Le Bris.
- **Profil recherché** : titulaire d'un master 2 ou diplôme d'ingénieur, avec de solides connaissances en apprentissage, géomatique, informatique ou bases de données.
- **Condition de nationalité** : Les candidats doivent être de nationalité française ou d'un pays de l'Union Européenne.
- **Date limite de candidature** : 6 Avril 2022
- Les dossiers de candidatures doivent comprendre **un CV, une lettre de motivation, des relevés de notes, ainsi que d'éventuelles lettres de recommandation** (bienvenues). Ils doivent être envoyés sous forme d'un **fichier pdf unique** aux adresses suivantes :  
*ana-maria.raimond(a)ign.fr ; laurence.jolivet(a)ign.fr ; arnaud.le-bris(a)ign.fr*

### Références bibliographiques

- Baudoux, L., Inglada, J., Mallet, C. (2021) Toward a country-scale land cover map without using images : a map translation approach. Remote Sensing 13(6), pp. 1060, 10.3390/rs13061060.
- Carvalho, M., Le Saux, B., Trouvé-Peloux, P., Almansa, A., Champagnat, F.(2019) Multi-task learning of height and semantics from aerial images. IEEE Geoscience and Remote Sensing Letters, 17(8):1391-1395.
- Gaetano, R., Ienco, D., Ose, K., Cresson, R. (2018) A two-branch CNN architecture for land cover classification of PAN and MS Imagery. Remote Sensing, 10(11):1746.
- Landrieu, L., Simonovsky, M. (2018) Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In Proceedings of IEEE CVPR, pp. 4558-4567.
- Lebedev, M.A., Vizilter, Y.V., Vygolov, O.V., Knyaz, V.A., Rubis, A.Y. (2018) Change detection in remote sensing images using Conditional Adversarial Networks. International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences, XLII(2):565-571.
- Liu, L., Olteanu-Raimond A-M, Jolivet L., le Bris A., See L, (2020) A data fusion-based framework to integrate multi-source VGI in an authoritative land use database, International Journal of Digital Earth.

- 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.
- 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.
- Papadomanolaki, M., Verma, S., Vakalopoulou, M., Gupta, S., Karantzalos, K. (2019) Detecting urban changes with recurrent neural networks from multitemporal Sentinel-2 data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'2019), Yokohama, Japan, 28 July - 2 August 2019.
- Peng, D., Zhang, M., Wanbing, G. (2019) End-to-End change detection for high resolution satellite images using improved UNet++. *Remote Sensing* 11:1382.
- Pomente, A., Picchiani, M., Del Frate, F. (2018) Sentinel-2 Change detection based on deep features. In Proceedings of the IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, July 2018; pp. 6859-6862.
- Stocker, O., Le Bris, A. (2020) Can SPOT-6/7 CNN semantic segmentation improve Sentinel-2 based land cover products? Sensor assessment and fusion. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol V-2, pp. 557-564, 10.5194/isprs-annals-V-2-2020-557-2020.