



Résumés des communications

journée d'étude
2 décembre 2021

campus des Grands Moulins / Université de Paris

Accès et utilisation des données en libre accès en géographie de la population, de la santé et des mobilités

Comité d'organisation

- Mickaël BLANCHET, CNFG, Université d'Angers
- Audrey BOCHATON, CIST, Université Paris-Nanterre
- Delphine BROUSMICHE, CIST, Université de Lille
- Yoann DOIGNON, CNFG, UC Louvain
- Julie FEN-CHONG, CIST, Université de Bourgogne
- Christophe IMBERT, CNFG-CIST, Université de Rouen Normandie
- Charlène LE NEINDRE, CIST, IRDES
- Françoise LUCCHINI, CIST, Université de Rouen

Contexte

L'*open data* (données en libre accès) désigne l'ensemble des données (sous différentes formes) mises en ligne et libres de droit. Comme le rappellent Ollion et Boellaert (2015), « si le label (*open data*) regroupe en théorie des données produites par une organisation et rendues globalement disponibles, ce sont généralement les données organisationnelles accessibles portées par des institutions publiques *via* internet que l'on désigne ainsi ». Localisées sur des plateformes collaboratives (*OpenStreetMap*) ou des plateformes gérées par des institutions publiques, les données en libre accès s'avèrent hétéroclites tant sur la forme (bases de données statistiques, photosatellite, zonage, outil de géolocalisation, cartes réalisables en ligne, fonds de carte...) que dans leur contenu (données démographiques, liste d'établissements médicaux, sportifs ou culturels, budget d'une administration, données relatives aux actes de santé ou aux bénéficiaires de prestations sociales, services de mobilité...). Polymorphes et territorialisées à différentes échelles, ces données délivrent des informations relatives :

- aux caractéristiques et aux pratiques (exemples : recours aux soins, mobilité résidence principale–lieu de travail, migrations résidentielles...) des populations,
- à la présence et à la densité de fonctions (paysagère, économique...), de services, de réseaux et d'équipements,
- aux activités des populations (bénéficiaires, fréquentation, recours...), aux fonctions, aux mobilités et aux flux.

Si les données en libre accès constituent pour les chercheur-es, les enseignant-es, les étudiant-es et les professionnel-le-s des sources précieuses d'informations, elles n'en restent pas moins l'objet de profondes limites. En premier lieu, ces données proviennent en majorité d'institutions publiques et/ou de plateformes collaboratives (Chignard, 2012), excluant au passage les organisations privées (dont les données explosent et recouvrent des informations relatives aux pratiques des populations). De plus, les données en libre accès représentent la surface, une part limitée des informations produites et utilisées par ces mêmes institutions et acteurs. Au même titre que pour les organisations privées, ne sont mises en ligne que les données susceptibles de ne pas entamer les capacités

et les politiques de ces mêmes acteurs. Enfin, ces données s'avèrent hétérogènes, tant dans leurs normes de collecte, leurs contenus que dans leurs accès. Cette hétérogénéité aboutit à un manque d'harmonisation géographique des données.

Enjeux, objectifs et sessions de la journée d'étude

Fort de ces constats, la commission population du CNFG et les axes Territoires et santé, Information territoriale locale, et Mobilités, identités et territoires du CIST organisent une journée d'étude consacrée à l'appréhension et à l'utilisation des données en libre accès en géographie de la population, de la santé et des mobilités. Le premier objectif de cette journée est d'offrir un panorama large de leurs modes d'appréhension, des enjeux de leur usage et des types de connaissance produites. Son second objectif est de parvenir à situer leur diffusion dans le phénomène plus large de développement des données numériques personnelles. S'adressant à des doctorant-es, jeunes docteur-es, chercheur-es et professionnel-le-s en géographie et sciences sociales intéressé-es par ces enjeux, la journée sera organisée en deux sessions :

- Une présentation de travaux en géographie de la population, de la santé et des mobilités. Lors de cette session, les présentations veilleront à resituer les apports et les limites empiriques des données en libre accès vis-à-vis de leurs démarches scientifiques.
- Des sous-groupes thématiques permettant aux participants d'échanger sur les sources (DVF, BDA accidents, *open data* locaux, SNDS, blablacar, twitter...) et applications (google data search, plateformes nationales et locales, géoportail, odomatrix...) en *open data* originales et propices à toute démarche en géographie de la population, de la santé ou des mobilités.

En complément de ces deux sessions, deux présentations de cadrage auront lieu : un cadrage juridique relatif aux mouvements législatifs et aux précautions d'utilisation des données numériques (en introduction) et un cadrage géographique relatif aux enjeux de l'étude des données en géographie (en conclusion).

Références bibliographiques¹

Boullier D., 2015, « Vie et mort des sciences sociales avec le *big data* », *Socio. La nouvelle revue des sciences sociales*, n° 4, p. 19-37 [doi.org/10.4000/socio.1259].

Cattaruzza A., 2019, *Géopolitique des données numériques. Pouvoir et conflits à l'heure du big data*, Paris, le Cavalier Bleu.

Chignard S., 2012, *L'open data, comprendre l'ouverture des données publiques*, Limoges, Fyp Éditions.

Ollion E., Boelaert J., 2015, « Au-delà des *big data* », *Sociologie*, 6(3) [journals.openedition.org/sociologie/2613].

Ollion E., 2015, « L'abondance et ses revers. *Big data*, *open data* et recherches sur les questions sociales », *Informations sociales*, n° 5, p. 70-79 [doi.org/10.3917/inso.191.0070].

N. Pinède, M. Noucher, F. Gourmelon, K. Soumagnac-Colin, 2017, « De l'*open data* à l'*open science* : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques », *Revue française des sciences de l'information et de la communication*, n° 11 [doi.org/10.4000/rfsic.3200].

1 Tous les liens du recueil ont été consultés courant octobre 2021.

Programme

- 9h00 Café d'accueil
- 9h30 Présentation de la journée d'étude par les organisateur-rices
- 9h45 **Conférence d'ouverture**
Valérie BAILLY (Institut du droit international des transports, Rouen) – Comment la législation impulse l'ouverture et le partage des données en libre accès ?
- 10h30 **Session 1. Géographie de la mobilité et de la population**
– Julie VALLÉE (CNRS/Géographie cités) – Diffuser et croiser des données libres en géographie : un retour d'expérience autour de l'outil *Mobiliscope*
– Boris MERICKSAY et Florent DEMORAES (Université de Rennes 2/ESO-Rennes) – Les données DVF en *open data* comme matériel de recherche pour les géographes : propositions méthodologiques à travers l'exemple de la Bretagne et la métropole rennaise
– Marianne GUÉROIS (Université de Paris/Géographie cités – Riate), Ronan YSEBAERT (Université de Paris/Riate) et Timothée GIRAUD (CNRS/Riate) – Apports et limites de la base *OpenStreetMap* pour l'analyse de l'accessibilité territoriale et de la localisation d'équipements
– Nicolas OVTRACHT (CNRS/LAET), Camille SIEPER (ENTPE/LAET), Boris MERICKSAY et Florent DEMORAES (Université de Rennes 2/ESO-Rennes), Sarah DUCHÉ et Kamila TABAKA (Université Grenoble Alpes/Pacte) – Construction d'une nomenclature unifiée des aménagements cyclables en France et en Colombie à partir de données en libre accès et institutionnelles : enjeux et perspectives
– Patrice TISSANDIER (Université Bordeaux Montaigne/Passage) et Jérémy BÉTHUNE (Université Paul-Valéry Montpellier III/master Transport, mobilités, réseaux) – Utilisation de données en libre accès visant à la création d'un outil d'aide à la décision pour l'implantation d'espaces d'enseignement à distance
- 12h30 Déjeuner

- 14h00 **Session 2. Géographie de la santé**
- Maxime THORIGNY (Université de Reims Champagne-Ardenne/Regards), Leila BOUAZZI et Stéphane SANCHEZ (Université de Reims Champagne-Ardenne/CUR2S), Alexandre DENOYER (CHU Robert Debré, Université de Reims Champagne-Ardenne/CardioVir) – Étude exposition aux pesticides et maladies neurodégénératives
 - Lara-Maria WAKIM, Florent OCCELLI, Damien CUNY, Caroline LANIER, Francis DOUAY et Annabelle DERAM (Université de Lille/LGCgE) – Maladie de Crohn et pollution des sols
 - Delphine BROUSMICHE, Caroline LANIER, Annabelle DERAM, Florent OCCELLI et Damien CUNY (Université de Lille/LGCgE), Michaël GENIN (Université de Lille/METRICS), Aline MEIRHAEGHE (Université de Lille/RID-AGE) – Développement d’indices composites spatialisés pour l’étude du lien entre environnement au sens élargi et survenue de pathologies chroniques : apports et limites de l’*open data*
 - Martin PAUMELLE, Florent OCCELLI, Caroline LANIER, Damien CUNY et Annabelle DERAM (Université de Lille/LGCgE) – Utilisation de données en libre accès pour caractériser le lien entre insuffisance rénale chronique terminale et environnement
 - Marie-Hélène HERMAND (Université Bordeaux Montaigne/MICA) – Les *open data* territorialisées à l’échelle transfrontalière : prescriptions, enjeux et réalités à l’aune d’un corpus de plateformes de surveillance épidémiologique
 - Jean-Marc MACÉ et Katerina KONONOVICH (CNAM/Lirsa) – Construction d’une source en *open data* sur le recours aux soins de première ligne
- 16h45 Pause café
- 17h00 **Conférence de clôture**
- Amaël CATTARUZZA (Institut français de géopolitique – président du CNFG) – Les conséquences sociospatiales et géopolitiques du développement des ressources en *open data*

Session 1.

Géographie de la mobilité et de la population

– Julie VALLÉE (CNRS/Géographie cités)

Diffuser et croiser des données libres en géographie : un retour d'expérience autour de l'outil *Mobiliscope*

– Boris MERICKSAY et Florent DEMORAES (Université de Rennes 2/ESO-Rennes)

Les données DVF en *open data* comme matériel de recherche pour les géographes : propositions méthodologiques à travers l'exemple de la Bretagne et la métropole rennaise

– Marianne GUÉROIS (Université de Paris/Géographie cités – Riate), Ronan YSEBAERT (Université de Paris/Riate) et Timothée GIRAUD (CNRS/Riate)

Apports et limites de la base *OpenStreetMap* pour l'analyse de l'accessibilité territoriale et de la localisation d'équipements

– Nicolas OVTRACHT (CNRS/LAET), Camille SIEPER (ENTPE/LAET), Boris MERICKSAY et Florent DEMORAES (Université de Rennes 2/ESO-Rennes), Sarah DUCHÉ et Kamila TABAKA (Université Grenoble Alpes/Pacte)

Construction d'une nomenclature unifiée des aménagements cyclables en France et en Colombie à partir de données en libre accès et institutionnelles : enjeux et perspectives

– Patrice TISSANDIER (Université Bordeaux Montaigne/Passage) et Jérémy BÉTHUNE (Université Paul-Valéry Montpellier III/master Transport, mobilités, réseaux)

Utilisation de données en libre accès visant à la création d'un outil d'aide à la décision pour l'implantation d'espaces d'enseignement à distance

Diffuser et croiser des données libres en géographie : un retour d'expérience autour de l'outil *Mobiliscope*

Julie VALLÉE

Mots clés

mobilité – inégalités – ségrégation – politiques publiques – géovisualisation

Cette communication propose un retour d'expérience sur la diffusion et le croisement de données libres en géographie, en lien avec le *Mobiliscope*¹, un outil libre de géovisualisation des variations de la population présente dans les territoires au cours de la journée.

1. Diffuser des données libres

Les variations spatiotemporelles de fréquentation des territoires peuvent être connues grâce aux données de téléphonie mobile qui localisent un très grand nombre d'individus avec une granularité spatiale et temporelle fine. L'accès à ces données est toutefois juridiquement contrôlé ce qui place les opérateurs téléphoniques en situation de pouvoir lorsqu'ils développent leurs solutions commerciales et non libres ou organisent des collaborations avec des équipes de recherche. L'idée est alors d'utiliser les vastes enquêtes publiques sur les déplacements quotidiens afin de développer le *Mobiliscope*, un outil qui donne à voir les fréquentations horaires des territoires et qui propose sous licence libre des données agrégées de présence par heure et par secteur (mais aussi par groupe social) pour 2 500 secteurs en France (10 000 communes). Le *Mobiliscope* est ainsi une alternative libre aux services payants et propriétaires issus des données de téléphonie mobile. Même si cet outil est développé à partir de données d'enquêtes sur les déplacements qui (à quelques exceptions près) ne sont pas librement libres, rien n'interdit en effet que les données agrégées de présence affichées dans l'outil, et les algorithmes ayant servi à les produire, soient librement mis à disposition.

2. Croiser des données libres

Dès lors que différentes données libres concernent la même zone géographique, il peut être intéressant de les croiser. Le *Mobiliscope* ne fait pas exception : les données agrégées sur les populations présentes par heure et par secteur peuvent ainsi être mises en relation avec d'autres données libres, que ce soit (1) la localisation spatiale des équipements² afin de calculer les densités d'équipements en fonction de la population présente (et non pas uniquement de la population résidente) ; (2) les périmètres d'intervention de l'action publique (les « quartiers prioritaires en politique de la ville »³, les

1 www.mobiliscope.cnrs.fr

2 www.data.gouv.fr/fr/datasets/base-permanente-des-equipements-1

3 www.data.gouv.fr/fr/datasets/quartiers-prioritaires-de-la-politique-de-la-ville-qpv

dispositifs « action cœur de ville »⁴ ou « petites villes de demain »⁵) afin de réaliser des typologiques de ces territoires d'intervention selon leur attractivité quotidienne et leur mixité fonctionnelle ; ou (3) les consommations locales d'électricité⁶ afin de les comparer aux volumes cumulés de populations présentes au cours de la journée et aux types d'activités réalisées. Ces trois exemples illustrent la diversité des pistes de recherche que le catalogue sans cesse croissant de données libres permet d'imaginer et d'explorer.

L'autrice

Julie VALLÉE

CNRS – UMR8504 Géographie-cités

julie.vallee@parisgeo.cnrs.fr

4 www.data.gouv.fr/fr/datasets/programme-action-coeur-de-ville

5 www.data.gouv.fr/fr/datasets/programme-petites-villes-de-demain

6 www.data.gouv.fr/fr/datasets/donnees-locales-denergie

Les données DVF en *open data* comme matériel de recherche pour les géographes : propositions méthodologiques à travers l'exemple de la Bretagne et la métropole rennaise

Boris MERICKSAY, Florent DEMORAES

Mots clés

demandes de valeurs foncières – *open data* – méthode – reproductibilité – outil libre – science ouverte

Les données immobilières ont connu ces dernières années de profondes transformations à la fois dans leurs modalités de production, de mise à disposition et d'utilisation. Longtemps restée affaire de spécialistes dans la mesure où ces informations n'étaient disponibles qu'au sein de bases complexes, parcellaires, peu accessibles voire payantes, la mise à disposition en *open data* de la base *Demandes de valeurs foncières* (DVF) produite par la direction générale des Finances publiques (DGFIP) en 2019, a ouvert de nouvelles perspectives. Celles-ci sont liées à la masse des données à traiter, aux possibilités de moduler les échelles d'analyse, aux opportunités offertes par la vaste palette d'outils *open source* et de méthodes aujourd'hui disponibles et à la combinaison de plusieurs modes de représentation. Ces perspectives concernent par ailleurs de nouveaux publics.

Les 17 millions de mutations foncières et immobilières collectées depuis 2014, renseignées dans la base DVF permettent désormais à des non spécialistes d'envisager l'observation et la compréhension des marchés immobiliers dans le temps et dans l'espace. Au même titre que les données issues de la statistique publique ou d'enquêtes, ces marqueurs détaillés des transactions immobilières, renseignés géographiquement de manière précise, permettent d'explorer, de quantifier et de spatialiser les structures des marchés de l'immobilier à différentes échelles. Toutefois, les bases DVF mises à disposition par les pouvoirs publics en accès libre et sous licence ouverte ne sont pas des données « prêtes à l'emploi » et diffèrent de celles accessibles aux ayant-droits de la DGFIP. Leur utilisation nécessite un temps de découverte et de compréhension qu'il ne faut pas sous-estimer. De plus, que ce soit pour leur préparation, leur analyse ou leur représentation (carto)graphique, les réutilisateurs doivent prendre en main des outils spécifiques et des méthodes appropriées qui requièrent un temps d'apprentissage.

C'est dans cette perspective que s'inscrit cette communication dont le but est de présenter une démarche documentée et reproductible à destination des géographes, des aménagistes et des urbanistes, pour appréhender spatialement les marchés immobiliers résidentiels à partir des données DVF en *open data*. Pour étayer la démonstration, nous développons à titre illustratif les exemples de la région Bretagne et de la métropole rennaise. Nous étudions plus spécifiquement les dynamiques spatiales des prix, des volumes d'achats / ventes et des sous-marchés immobiliers suivant une approche multiscale. L'ambition n'est pas de présenter un modèle expliquant les structures spatiales des prix de

l'immobilier, ni une monographie des marchés bretons, mais plutôt de fournir des pistes méthodologiques issues de la statistique descriptive, de l'analyse spatiale et des techniques cartographiques qui peuvent être répliquées sur d'autres zones d'étude.

L'objectif de la communication est de (1) formaliser la préparation des données DVF disponibles en open data (apurement, filtrage, restructuration, enrichissement) ; (2) les analyser sous différents angles et suivant plusieurs échelles (agrégation, classification) ; (3) les restituer (carto)graphiquement au moyen de plusieurs types de rendus (graphiques, cartes choroplèthes, cartes lissées, carroyage, cartogramme) et (4) discuter des avantages et des inconvénients de ces nouvelles données pour les géographes dans la compréhension des dynamiques territoriales.

Les auteurs

Boris Mericskay

Université Rennes 2 – UMR6590 Espaces et sociétés (ESO-Rennes)

boris.mericskay@univ-rennes2.fr

Florent Demoraes

Université Rennes 2 – UMR6590 Espaces et sociétés (ESO-Rennes)

florent.demoraes@univ-rennes2.fr

Apports et limites de la base *OpenStreetMap* pour l'analyse de l'accessibilité territoriale et de la localisation d'équipements

Marianne GUÉROIS, Timothée GIRAUD, Ronan YSEBAERT

Mots clés

OpenStreetMap – réseaux routiers – Points of Interest – comparaisons internationales

La base contributive *OpenStreetMap* (OSM) peut être exploitée en tant que base de données décrivant des objets localisés de diverses natures, depuis les réseaux routiers jusqu'aux équipements commerciaux et de services. Cette information géographique, qui constitue avant tout une ressource pour un grand nombre d'applications en ligne, est aussi utilisée dans un nombre croissant de travaux de recherche, que ce soit pour analyser la dispersion de la population à des échelles fines, identifier des lieux d'activités et de sociabilité ou encore, à des échelles plus vastes, comparer l'équipement des régions européennes en services à la population. Face à ces nouveaux usages, de nombreux travaux ont interrogés la qualité de données produites selon un mode décentralisé et non supervisé, et montré que ces informations atteignaient parfois des niveaux de qualité proches des bases conventionnelles, notamment pour les données de réseaux routiers.

Dans cette présentation, nous proposons un retour d'expérience sur l'utilisation de la base OSM dans le cadre de plusieurs études réalisées par l'UMS Riate, qui s'appuient d'un côté sur des mesures d'accessibilité territoriale, de l'autre sur l'analyse de la localisation d'équipements. Bien que provenant de la même source, les enjeux propres à l'analyse du potentiel de ces deux types de données diffèrent sensiblement. Dans le cas des réseaux routiers, nous exploitons la base OSM pour construire des mesures d'accessibilité à l'aide d'un outil issu de l'algorithme OSRM, implémenté dans le langage R (package *osrm*), qui permet de construire des distanciers. L'enjeu porte principalement sur les conditions d'extraction de l'information et sur le choix de l'algorithme optimal. Dans le cas des données d'équipement, la question de la qualité de l'information est centrale et préalable à toute analyse thématique. Dans l'exemple présenté (organisation spatiale des équipements commerciaux dans la Grande Région – espace de coopération transfrontalier entre France, Allemagne, Luxembourg et Belgique), il s'agissait de savoir dans quelle mesure le recours aux données d'OSM permettait de compenser l'absence d'une base européenne harmonisée sur les équipements. Du côté français, l'existence de la base permanente des équipements (BPE, Insee) a offert une opportunité unique pour évaluer la complétude d'OSM à une échelle régionale et souligné l'exploitation prudente qui doit être faite d'OSM dans ce cadre : l'information est plutôt fiable dans les centres-villes densément peuplés, plus aléatoire dans les périphéries urbaines et les zones rurales hors influence urbaine. Cette hétérogénéité spatiale entre différents types d'espace urbains / ruraux se double d'une importante hétérogénéité de contribution entre les pays de la zone d'étude.

Références bibliographiques

Bakillah M., Liang S. H. L., Mobasheri A., Jokar Arsanjani J., Zipf A., 2014, « Fine Resolution Population Mapping Using OpenStreetMap Points-of-Interest », *International Journal of Geographical Information Science*, 28(9), p. 1940-1963 [doi.org/10.1080/13658816.2014.909045].

Bawa-Cavia A., 2011, « Sensing the Urban: Using Location-Based Social Network Data in Urban Analysis », communication à l'atelier « Pervasive Urban Applications (PURBA) », 9^e colloque international « Pervasive Computing », San Francisco, USA.

ESPON, 2017, *PROFECY. Inner Peripheries: National territories facing challenges of access to basic services of general interest* [www.espon.eu/inner-peripheries].

Giraud T., 2019, « OpenStreetMap pour la recherche sur l'accessibilité », présentation à la journée d'étude « Accessibilité spatiale aux soins », CIST, Paris, 12 juin 2019 [cist.cnrs.fr/accessibilite-spatiale-aux-soins-eclairages-theoriques-et-methodologiques].

Guérois M., Ysebaert R., Giraud T., Maranget B., 2020, *Apports des données OpenStreetMap pour l'analyse des centralités commerciales dans les espaces transfrontaliers*, rapport pour l'Agence nationale de la cohésion des territoires (ANCT) [rcarto.gitpages.huma-num.fr/centralite/report/report.pdf].

Haklay M., 2010, « How Good is Volunteered Geographic Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets », *Environ. Plan. B*, n° 37, p. 682-703 [doi.org/10.1068/b35097].

Hayat F., 2019, *Production des biens communs numériques et usages cartographiques*, thèse de doctorat, Université Paris Diderot.

Hombiat A., Villanova-Oliver M., Gensel J., 2016, « Un métamodèle pour structurer la folksonomie d'OSM en une nouvelle ontologie », *Revue internationale de géomatique*, n° 4/2016, p. 425-444.

Jonietz D., Zipf A., 2016, « Defining Fitness-for-Use for Crowdsourced Points of Interest (POI) », *ISPRS International Journal of Geo-Information*, 5(9), p. 149-161 [doi.org/10.3390/ijgi5090149].

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(3), p. 80-109 [doi.org/10.3390/ijgi6030080].

Viry M., Giraud T., Guérois M., Ysebaert R., Lambert N., Feredj A., 2016, *Généralités liées à OpenStreetMap et à la complétude des données*, CGET–UMS Riate.

Les auteur·es

Marianne GUÉROIS

Université de Paris – UMR8504 Géographie-cités / UMS2412 Réseau interdisciplinaire pour l'aménagement, l'observation et la cohésion des territoires européens (Riate)
marianne.guerois@u-paris.fr

Timothée GIRAUD

CNRS – UMS2412 Réseau interdisciplinaire pour l'aménagement, l'observation et la cohésion des territoires européens (Riate)
timothee.giraud@cnrs.fr

Ronan YSEBAERT

Université de Paris – UMS2412 Réseau interdisciplinaire pour l'aménagement, l'observation et la cohésion des territoires européens (Riate)
ronan.ysebaert@cnrs.fr

Construction d'une nomenclature unifiée des aménagements cyclables en France et en Colombie à partir de données institutionnelles ou en libre accès : enjeux et perspectives

Nicolas OVTRACHT, Camille SIEPER, Boris MERICKSKAY,
Florent DEMORAES, Sarah DUCHÉ, Kamila TABAKA

Mots clés

aménagements cyclables – données spatiales en libre accès – données *OpenStreetMap* – « coronapistes » – typologie des pistes cyclables

Problématique

Ce travail s'inscrit dans le programme de recherche Vélotactique (programme d'un an et demi financé par l'Agence nationale française de la recherche), dont un des objectifs est de collecter, structurer et analyser des données permettant aux praticiens comme aux chercheurs, d'apporter des connaissances sur les réseaux cyclables permanents et sur les pistes temporaires ou « coronapistes » mises en place suite à la crise de la Covid-19.

Ce projet vise à comprendre les impacts de la crise sanitaire sur les aménagements cyclables dans plusieurs métropoles en France, en Suisse, au Québec et en Colombie. Dans cette optique, les méthodes et outils de la géomatique sont particulièrement pertinents pour produire des indicateurs spatialisés, comparer les territoires et éclairer la décision publique.

Données et méthodologie

Le défi a été de constituer des bases de données spatiales harmonisées sur les aménagements cyclables et de mettre en place des méthodes répliquables sur d'autres terrains. À cette fin, nous avons utilisé les données des métropoles (plateforme *open source* et contacts avec des techniciens locaux) et les données *OpenStreetMap*. Nous avons d'abord mis en œuvre la démarche sur un échantillon de métropoles françaises de taille variée, à savoir Grenoble, Lyon, Montpellier et Rennes, en tenant compte des spécifications des sources de données. Nous avons utilisé l'hybridation des données pour compléter la composante géométrique des aménagements et fait appel à des données thématiques pour établir une typologie des pistes cyclables suivant les normes du Cerema (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) et ETALAB¹. Cela nous a permis de calculer des statistiques descriptives telles que le linéaire de voirie par type d'aménagement pour comparer les territoires

1 ETALAB est un département de la direction interministérielle du numérique (DINUM), dont les missions et l'organisation sont fixées par le décret du 30 octobre 2019. Faisant office de *Chief Data Officer* de l'État (au titre des missions de l'administrateur général des données, des algorithmes et des codes sources), il coordonne la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée ; la politique d'ouverture et de partage des données publiques (*open data*) ; et les actions des administrations de l'État et leur apporte son appui pour faciliter la diffusion et la réutilisation de leurs informations publiques (www.etalab.gouv.fr).

français. Dans le cadre de ce projet, nous avons dans un deuxième temps transposé cette typologie sur Bogotá (Colombie) afin de vérifier la répliquabilité de la méthode.

Résultats

Les résultats ont permis de comparer les 4 métropoles françaises étudiées et Bogotá et, ainsi, d'analyser la structure et l'envergure des pistes cyclables sur les différents territoires, notamment le déploiement des dites « coronapistes ». À ce stade, les coronapistes des métropoles françaises constituent une part modérée du réseau des pistes cyclables (moins de 8 % de linéaires supplémentaires). Par contre, dans le district de Bogotá, l'accroissement du linéaire du réseau cyclable a été de l'ordre de 14 % dans les mois suivant le premier déconfinement 2020, même si aujourd'hui une grande partie du réseau temporaire a été démantelée.

Les auteurs

Florent DEMORAES

Université Rennes 2 – UMR6590 Espaces et Sociétés (ESO-Rennes)

florent.demoraes@univ-rennes2.fr

Sarah DUCHÉ

Université Grenoble Alpes – UMR5194 PACTE

sarah.duche@univ-grenoble-alpes.fr

Boris MERICKSKAY

Université Rennes 2 – UMR6590 Espaces et Sociétés (ESO-Rennes)

boris.merickskay@univ-rennes2.fr

Nicolas OVTRACHT

CNRS – UMR5593 Laboratoire aménagement, économie, transport (LAET)

nicolas.ovtracht@cnrs.fr

Camille SIEPER

ENTPE – UMR5593 Laboratoire aménagement, économie, transport (LAET)

camille.sieper@entpe.fr

Kamila TABAKA

Université Grenoble Alpes – UMR5194 PACTE

kamila.tabaka@univ-grenoble-alpes.fr

Utilisation de données en libre accès visant à la création d'un outil d'aide à la décision pour l'implantation d'espaces d'enseignement à distance

Patrice TISSANDIER, Jérémy BÉTHUNE

Mots clés

enseignement à distance – outil d'aide à la décision – données en libre accès – données cartographiques – données alphanumériques

Existant depuis les années 1960, l'enseignement à distance, défini comme « un enseignement–apprentissage dont la médiatisation permet de rapprocher le savoir de l'apprenant, alors qu'enseignant et apprenant sont éloignés l'un de l'autre » (CLIFAD, 2007), a connu plusieurs phases d'essor au XX^e siècle, conjointement à la progression de la micro-informatique et des télécommunications (Hantem, 2020), puis une large diffusion en France dans notre étude depuis le printemps 2020 et les périodes successives de confinement et d'enseignement à distance imposées par le MESR.

En parallèle à cette crise sanitaire, le MESR accompagne la transition des universités françaises dans le développement de l'enseignement à distance, comme l'illustrent les 89 lieux labellisés « campus connecté », dispositif qui a vocation à perdurer et dont l'objectif est, entre autre, de s'implanter en zone rurale afin d'améliorer l'accessibilité à l'enseignement supérieur et sa qualité pour un public qui ne bénéficie pas obligatoirement d'un accès à une connexion internet fiable, d'un logement adapté ou d'un matériel informatique suffisant.

Si les objectifs sont louables, de récentes études ont permis de souligner les limites d'un tel système, qui s'expriment notamment par une perte du lien social, un sentiment d'isolement ou encore une impression de dégradation des interactions pédagogiques (Granjon, 2021).

Dans ce cadre, à l'image des évolutions récentes que connaissent les espaces de *coworking* vers des tiers-lieux qui ne sont plus uniquement des espaces de travail exclusif, il semble important de pouvoir proposer des espaces adaptés aux étudiants qui souhaitent bénéficier de conditions d'apprentissage de qualité.

Ainsi, en s'appuyant sur des données en libre accès provenant de sources diverses, nous proposons de développer un outil d'aide à la décision afin d'optimiser l'implantation d'espaces de *colearning*. Un travail exploratoire préliminaire réalisé sur le territoire de la région Nouvelle Aquitaine a été mené au printemps et à l'été 2021, qui a permis de produire un inventaire des données en libre accès en mettant en relief l'obligation de mobiliser des données produites à différentes échelles spatiales. À l'échelle macro, les données

issues de producteurs nationaux (Insee, IGN ou IFEN par exemple) ou de sites collaboratifs (OSM) et, à l'échelle méso, celles produites par des services régionaux (plateforme d'échange PIGMA ou GIP ATGERI par exemple qui regroupe l'État, l'ONF, le conseil régional Nouvelle-Aquitaine, ou certains services départementaux notamment) font preuve d'une certaine homogénéité dans leur production et dans leur diffusion, facilitant leur intégration dans la base de données. *A contrario*, à l'échelle micro (départementale et infra-départementale), plusieurs problématiques apparaissent :

- de qualité : à titre d'exemple, les réseaux de pistes cyclables ne sont constitués que de tronçons non reliés entre eux ;
- d'hétérogénéité du contenu : en fonction de l'échelle, différents niveaux de précision sont source de blocages ;
- d'interopérabilité : en fonction de l'organisme producteur, pour une donnée sur un thème identique ;
- d'existence : certaines données produites à l'échelle micro (cas fréquent des intercommunalités) sont très éparpillées et ne couvrent pas l'ensemble du territoire d'étude.

Il apparaît ainsi clairement, à l'issue de cette phase exploratoire de l'étude, que, en Nouvelle Aquitaine, l'identification de sites potentiels d'implantation d'espaces de *colearning* à une échelle fine doit faire face à ces problématiques qui posent, par ailleurs, la question de la potentielle reproductibilité de l'outil d'aide à la décision sur d'autres territoires régionaux.

Références bibliographiques

CLIFAD (Comité de liaison en formation à distance), 2007, *La formation à distance vue de près*.

Granjon Y., 2021, « La perception de l'enseignement à distance par les étudiants en situation de confinement : premières données », *Distances et médiations des savoirs*, 33 | 2021 [doi.org/10.4000/dms.6166].

Hantem A., 2020, « Les conditions de l'enseignement à distance pendant le confinement dû au COVID19. Cas de l'enseignement supérieur au Maroc » [hal.archives-ouvertes.fr/hal-02883214/document].

Les auteur·es

Patrice TISSANDIER

Université Bordeaux Montaigne – UMR Passages

patrice.tissandier@u-bordeaux-montaigne.fr

Jérémy BÉTHUNE

Université Paul-Valéry Montpellier III – Master Transport, mobilités, réseaux

jeremy.bethune@etu.univ-montp3.fr

Session 2. Géographie de la santé

– Maxime THORIGNY (Université de Reims Champagne-Ardenne/Regards),
Leila BOUAZZI et Stéphane SANCHEZ (Université de Reims Champagne-Ardenne/CUR2S),
Alexandre DENOYER (CHU Robert Debré, Université de Reims Champagne-Ardenne/
CardioVir)

Étude exposition aux pesticides et maladies neurodégénératives

– Lara-Maria WAKIM, Florent OCCELLI, Damien CUNY, Caroline LANIER, Francis DOUAY et
Annabelle DERAM (Université de Lille/LGCgE)

Maladie de Crohn et pollution des sols

– Delphine BROUSMICHE, Caroline LANIER, Annabelle DERAM, Florent OCCELLI et
Damien CUNY (Université de Lille/LGCgE), Michaël GENIN (Université de Lille/METRICS),
Aline MEIRHAEGHE (Université de Lille/RID-AGE)

Développement d'indices composites spatialisés pour l'étude du lien entre environne-
ment au sens élargi et survenue de pathologies chroniques : apports et limites de l'*open
data*

– Martin PAUMELLE, Florent OCCELLI, Caroline LANIER, Damien CUNY et Annabelle DERAM
(Université de Lille/LGCgE)

Utilisation de données en libre accès pour caractériser le lien entre insuffisance rénale
chronique terminale et environnement

– Marie-Hélène HERMAND (Université Bordeaux Montaigne/MICA)

Les *open data* territorialisées à l'échelle transfrontalière : prescriptions, enjeux et réalités
à l'aune d'un corpus de plateformes de surveillance épidémiologique

– Jean-Marc MACÉ (CNAM/Lirsa) et Katerina KONONOVICH (HESAM/Lirsa)

Construction d'une source en *open data* sur le recours aux soins de première ligne

Étude exposition aux pesticides et maladies neurodégénératives

Maxime THORIGNY, Leila BOUAZZI, Alexandre DENOYER, Stéphane SANCHEZ

Mots clés

maladies neurodégénératives – pesticide – exposition environnementale

Nous proposons un retour d'expérience de l'utilisation de bases de données en libre accès dans le cadre d'un projet scientifique visant à étudier l'association entre l'exposition aux pesticides et le taux de recours d'une maladie neurodégénérative¹. Cette recherche exploratoire vise à tester différents modèles d'exposition de la population aux pesticides.

Afin de calculer une dose d'exposition aux pesticides, nous avons utilisé la base de données *Achats de pesticides par code postal* produite par eaufrance². Cela a permis d'améliorer les méthodologies existantes qui codent l'exposition par « exposé » ou « non-exposé », et donc d'avoir une donnée plus précise et ainsi rendre ces travaux exploratoires originaux en travaillant à une maille territoriale fine. Les taux de recours ont été calculés à partir du programme de médicalisation des systèmes d'information (PMSI). L'étude a été réalisée aux niveaux départemental et communal. Les communes des bases de données du PMSI sont un regroupement de codes postaux. Pour les données d'achats de pesticides, une correspondance a été effectuée avec les codes Insee pour permettre une jointure entre les deux fichiers.

Cette étude est conduite en trois phases.

Dans une première phase, le ratio entre les quantités d'achats de pesticide par commune et sa surface a été calculé et découpé en classe. Les associations entre la quantité / hectare et le taux de recours ont été évalués au niveau communal. Nous avons ensuite testé deux modèles plus réalistes en agrégeant les données d'exposition des communes voisines, puis des voisines des voisines. Au niveau communal, une relation significative a été obtenue. Néanmoins, nous faisons l'hypothèse que la relation dose–effet peut être améliorée en prenant en considération le type de culture et son emprise exacte.

Suite à cela, dans une deuxième phase, nous souhaitons intégrer à l'exposition la proximité et le type de cultures. Ces données issues du *Registre parcellaire graphique 2.0*³

1 Les données de cette étude sont en cours de publications aussi nous ne pouvons pas dévoiler à ce stade de quelle maladie neurodégénérative il s'agit. Cela ne retire rien à l'exercice et nous présenterons les résultats sans y faire référence.

2 Données disponibles en ligne [www.data.gouv.fr/fr/datasets/achats-de-pesticides-par-code-postal/#resources].

3 Données disponibles en ligne [www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-cultureux-et-leur-groupe-de-cultures-majoritaire].

nous aideront à affiner l'exposition de la population aux pesticides à trois échelles : communale, communes voisines et voisines des voisines.

Enfin dans une troisième phase, nous testerons un autre modèle d'exposition à partir des données précédentes en utilisant des *buffers* de tailles variables afin de nous passer des échelles communales qui induisent des biais de différentes natures (surface variable, effet de bord).

Cependant cette approche reste dépendante, d'une part, de la qualité des données relatives d'exposition aux pesticides utilisés qui constituent des données de quantité d'achats et non d'utilisation, d'autre part, de la maille territoriale imposée (code postal), qui implique des regroupements de communes donc une perte de précision spatiale. De plus, ces données communales ont une tendance historique au regroupement, ce qui diminue le nombre d'individus et donc augmente la perte de précision.

Les auteur-es

Maxime THORIGNY

Université de Reims Champagne-Ardenne – EA6292 Recherches en économie gestion agroressources durabilité santé (Regards)

maxime.thorigny@gmail.com

Leila BOUAZZI

Université de Reims Champagne-Ardenne – Comité universitaire de ressources pour la recherche en santé (CUR2S)

leila.bouazzi@univ-reims.fr

Alexandre DENOYER

CHU Robert Debré, Reims – Service d'ophtalmologie

Université de Reims Champagne-Ardenne – EA4684 Infections cardiovasculaires virales et inflammation en pathologie humaine (CardioVir)

alexandre.denoyer@gmail.com

Stéphane SANCHEZ

Université de Reims Champagne-Ardenne – Comité universitaire de ressources pour la recherche en santé (CUR2S)

Hôpital de Troyes – Département d'information médicale

stephane.sanchez@hcs-sante.fr

Maladie de Crohn et pollution des sols

Lara-Maria WAKIM, Florent OCCELLI, Damien CUNY,
Caroline LANIER, Francis DOUAY, Annabelle DERAM

Mots clés

maladie de Crohn – pollution des sols – caractérisation des sols

La maladie de Crohn (MC) est une maladie chronique inflammatoire qui affecte le tractus gastro-intestinal. Si son étiologie n'est pas encore connue, l'influence de facteurs environnementaux est largement soutenue. Dans le cadre de notre étude, nous nous sommes attachés à la caractérisation des sols basée sur l'analyse spatiale des clusters de la maladie, qui ont été identifiés dans le nord de la France, grâce au registre EPIMAD¹, le plus grand registre des maladies inflammatoires chroniques de l'intestin comprenant la maladie de Crohn et la rectocolite hémorragique dans le nord-ouest de la France.

Cette étude vise à caractériser les sols à partir des données existantes. Cependant, de nombreuses limitations sont apparues, notamment la faible disponibilité des données à grande échelle et la non-accessibilité de certaines, sans parler d'un certain manque de fiabilité.

La distinction des différentes occupations de sols a été effectuée en utilisant la base de données en libre accès CORINE Land Cover² afin de cartographier et d'analyser l'occupation du sol dans les clusters. La comparaison entre les clusters a montré une dominance des territoires artificiels dans les clusters de sous-incidence, tandis que les terres agricoles sont prépondérantes dans les clusters de sur-incidence.

La densité d'industrialisation ainsi que la classification des activités (code NAF) ont également été étudiées à partir de l'inventaire historique des anciens sites industriels et activités de service (BASIAS³), qui renseigne sur la localisation des sites industriels potentiellement pollués et la nature des activités selon la classification NAF. Cette base de données présente des problèmes de fiabilité à cause de diverses limitations dans l'enregistrement de données dont certaines déjà évoquées (données manquantes, sites non cartographiés, inégalité d'enregistrement entre départements) et de manques d'information sur le type de polluants et le degré de contamination. Ces limitations nous ont poussés à utiliser cette donnée comme un facteur d'influence plutôt que d'exclusion.

1 www.observatoire-crohn-rch.fr/epimad-le-plus-grand-registre-de-malades-au-monde

2 www.data.gouv.fr/en/datasets/corine-land-cover-occupation-des-sols-en-france

3 www.georisques.gouv.fr/risques/basias/donnees#

Le Réseau de mesure de la qualité des sols (RMQS⁴) est un outil de surveillance de la qualité des sols développé en France. Seuls les sols agricoles et forestiers ont été échantillonnés et cartographiés sur une grille carrée de 16 km de côté, plus large pour notre étude dans la mesure où certains clusters ne comportent pas de sites d'échantillonnage alors que d'autres sont principalement des zones urbaines. Malgré la disponibilité des cartes interpolées, les données issues du RMQS ne reflètent pas la réalité des sols des clusters.

Face à l'ensemble de ces limitations, une campagne d'échantillonnage devrait permettre d'identifier plus clairement ces contaminants et de comparer les caractéristiques du sol entre les clusters.

Les auteurs

Lara-Maria WAKIM

Université de Lille – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
thèse financée par PEARL de l'I-SITE
laramaria.wakim@univ-lille.fr

Florent OCCELLI

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
florent.occelli@univ-lille.fr

Damien CUNY

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
damien.cuny@univ-lille.fr

Caroline LANIER

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
caroline.lanier@univ-lille.fr

Francis DOUAY

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
caroline.lanier@univ-lille.fr

Annabelle DERAM

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
annabelle.deram@univ-lille.fr

4 Les données de la campagne initiale sont disponibles en accès libre sur le site de GISSol [www.gissol.fr/le-gis/programmes/rmqs-34] pour chaque élément à part et téléchargeable via le GeoServer de l'INRAE [agroenvgeo.data.inra.fr/geonetwork/srv/fre/catalog.search#].

Développement d'indices composites spatialisés pour l'étude du lien entre environnement au sens élargi et survenue de pathologies chroniques : apports et limites de l'*open data*

Delphine BROUSMICHE, Caroline LANIER, Annabelle DERAM, Florent OCCELLI,
Michaël GENIN, Aline MEIRHAEGHE, Damien CUNY

Mots clés

infarctus du myocarde – inégalités territoriales – interopérabilité – déterminants de santé – multidimensionnalité des phénomènes

L'état de santé des populations est fortement influencé par des interactions complexes entre l'ensemble des déterminants de l'environnement, ceux-ci étant interconnectés et évoluant dans le temps et l'espace. Au regard de la diversité des facteurs impliqués en santé environnement ayant un impact positif ou négatif sur la santé, des approches multidimensionnelles s'avèrent nécessaires. Une des méthodes reconnues pour traiter une problématique complexe repose sur le développement d'indices composites permettant de rapprocher des domaines habituellement traités séparément. La transition d'un système complexe vers une mesure simple fait face à de nombreux défis tels que la qualité, la disponibilité et/ou l'accessibilité des données, l'unité spatiale choisie pour le développement de l'analyse conditionnant par ailleurs certains de ces défis (e.g. disponibilité, exhaustivité).

Face à ce constat, une approche associant des indices composites spatialisés de vulnérabilité (retranscrivant les contraintes et nuisances péjoratives auxquelles sont soumises les populations) et de résilience (définissant la capacité du territoire à faire face aux inégalités de santé) est développée pour définir les caractéristiques territoriales des îlots regroupés pour l'information statistique (IRIS) de la métropole européenne de Lille. Cette unité spatiale désigne le découpage administratif de base en matière de diffusion des données infra-communales. Les choix de l'unité spatiale et du périmètre d'analyse ont été conditionnés par la collaboration avec le registre MONICA¹, qui recense de manière exhaustive les pathologies coronariennes sur certains territoires. Afin de proposer une méthodologie transposable et accessible, notre méthodologie de développement d'indices composites s'appuie sur de l'*open data* pour la collecte de données visant à caractériser les déterminants de la santé environnement. Les organismes officiels (Insee, DREES...) ont été privilégiés pour garantir la qualité des données à intégrer aux indices composites, (i) directement ou (ii) après calculs (usuels, issus d'une adaptation transdisciplinaire, inspirés de précédents travaux...). Les indices seront mis en relation avec le taux d'incidence standardisé afin d'analyser l'influence des potentiels de vulnérabilité et/ou de résilience sur la survenue de pathologies coronariennes.

Toutefois, la multiplicité des fournisseurs de données soulève la question de leur interopérabilité. Des modifications des limites administratives (création, fusion...) sont

1 cprv.pagesperso-orange.fr/monica.htm

possibles, ce qui engendre des différences de granularité spatiale entre les données téléchargées. De plus, les fournisseurs de données ne proposent pas nécessairement des mises à jour annuelles, introduisant une différence de temporalité dans les données intégrées aux indices composites. Ces nombreuses sources de données ont l'avantage de permettre de caractériser les multifacettes des phénomènes impliqués en santé environnement, i.e. la santé, l'environnement, la politique, les services (incluant l'habitat), l'économie et le social (dont l'éducation).

Les auteur-es

Delphine BROUSMICHE

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)

delphine.brousmiche.etu@univ-lille.fr

Caroline LANIER

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)

Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)

caroline.lanier@univ-lille.fr

Annabelle DERAM

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)

Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)

annabelle.deram@univ-lille.fr

Florent OCCELLI

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)

Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)

florent.occelli@univ-lille.fr

Michaël GENIN

Univ. Lille – ULR2694 Évaluation des technologies de santé et des pratiques médicales (METRICS)

CHU de Lille – Unité de méthodologie biostatistique et data-management

michael.genin@univ-lille.fr

Aline MEIRHAEGHE

Univ. Lille, CHU Lille, Institut Pasteur de Lille, INSERM – UMR1167 Risk Factors and Molecular Determinants of Aging-Related Diseases (RID-AGE)

aline.meirhaeghe@pasteur-lille.fr

Damien CUNY

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)

Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)

damien.cuny@univ-lille.fr

Utilisation de données en libre accès pour caractériser le lien entre insuffisance rénale chronique terminale et environnement

Martin PAUMELLE, Florent OCCELLI, Caroline LANIER,
Damien CUNY, Annabelle DERAM

Mots clés

environnement – santé – insuffisance rénale – clusters spatiaux – étiologie

Certaines pathologies chroniques présentent une étiologie multifactorielle encore en partie indéterminée pour laquelle un lien avec l'environnement est suspecté. C'est le cas de l'insuffisance rénale chronique terminale (IRCT) pour laquelle un registre de santé (Néphronor) est tenu à jour depuis 2005 dans la région Hauts-de-France. Ces données recueillies par les professionnels de santé ont permis de calculer l'incidence de l'IRCT dans chaque commune et ainsi d'étudier son hétérogénéité dans l'espace. Grâce à des approches épidémiologiques géographiques, des clusters de sur-incidence et de sous-incidence¹ ont été détectés.

L'objectif de ces travaux est de contribuer à identifier de nouvelles pistes étiologiques en lien avec l'environnement. La démarche générale consiste à mettre en lien les données de santé avec des données environnementales² disponibles en libre accès. Après avoir défini les limites du sujet et les critères d'inclusion de nouvelles variables, une base de données environnementale a été construite. Un état des lieux a permis de retenir des données en *open data* couvrant des thématiques aussi diverses que l'occupation des sols, la pollution des milieux, la défaveur sociale, l'accès aux soins, etc.

Pour exploiter cette base de données et formuler des hypothèses sur l'étiologie de l'IRCT, deux méthodes de caractérisation de l'environnement ont été mises en œuvre. Une première consiste à utiliser des indices spatiaux composites : des indices de vulnérabilité et de résilience ont permis de quantifier le cumul d'effets délétères ou d'effets bénéfiques de l'environnement sur la santé, afin de tester le poids global de l'environnement sur la répartition spatiale d'incidence de l'IRCT. La seconde approche est une méthode de partitionnement du territoire qui permet de construire une typologie des communes en regroupant celles qui présentent un profil environnemental similaire. La méthode repose sur deux étapes successives : une analyse en composantes principales suivie d'une classification ascendante hiérarchique, afin de déterminer si certains profils environnementaux sont surreprésentés ou sous-représentés au sein des clusters.

1 Un cluster de sur-incidence est une zone géographique où l'incidence de la maladie est anormalement élevée par rapport à la zone d'étude. Au contraire, un cluster de sous-incidence présente une incidence anormalement faible.

2 L'environnement est compris au sens de milieu de vie des populations (physique, social, économique) en excluant ce qui relève du comportement individuel.

Dans ce contexte de recherche, l'avantage indéniable de l'*open data* est la possibilité de générer des hypothèses étiologiques à moindre coût. Avant d'envisager des études épidémiologiques individuelles beaucoup plus coûteuses, il permet de tester ces hypothèses à l'échelle du territoire grâce à une approche géographique et d'identifier des pistes qui pourront ensuite être mises à l'épreuve à l'échelle individuelle. Cependant son utilisation présente certaines limites dont une première assez évidente : la dépendance du travail de recherche vis-à-vis des données disponibles, qui doivent être fiables, pertinentes, exhaustives et disponibles à l'échelle considérée. D'autre part, il est crucial de garder un esprit critique sur la qualité des données utilisées et sur les résultats qui en découlent. Par exemple, pour étudier la contamination des milieux, il est assez rare de disposer de données directes mesurées dans l'environnement. Cela impose d'utiliser des proxys, c'est-à-dire d'évaluer indirectement le phénomène en considérant par exemple les sources d'émissions (avec toutes les limites que cela comporte). Un autre point à garder à l'esprit concerne la temporalité des données. Les clusters IRCT ont par exemple été détectés sur la période 2005-2018. Or il est généralement compliqué de remonter aussi loin dans le temps avec l'*open data*, sachant qu'idéalement il faudrait disposer de données encore plus anciennes pour prendre en compte le temps de latence potentiel entre une exposition à un environnement dégradé et le déclenchement d'une pathologie.

Les auteur·es

Martin PAUMELLE

Université de Lille – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
martin.paumelle.etu@univ-lille.fr

Florent OCCELLI

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
florent.occelli@univ-lille.fr

Caroline LANIER

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
caroline.lanier@univ-lille.fr

Damien CUNY

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
damien.cuny@univ-lille.fr

Annabelle DERAM

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)
annabelle.deram@univ-lille.fr

Les *open data* territorialisées à l'échelle transfrontalière : prescriptions, enjeux et réalités à l'aune d'un corpus de plateformes de surveillance épidémiologique

Marie-Hélène HERMAND

Mots clés

information territoriale – plateformes de santé – données épidémiologiques – transfert transfrontalier de données – observatoires de santé

Nous nous intéressons aux *open data* territorialisées à l'échelle transfrontalière et plus précisément aux plateformes de surveillance épidémiologique élaborées en contextes transfrontaliers dans différentes aires mondiales. Ces données renvoient à des enjeux de coopération (sanitaires, politiques, techniques, sociaux) formulés par des organisations internationales, des gouvernements nationaux, des instituts de recherche, des citoyens et des médias. Après avoir défini l'expression « surveillance épidémiologique » et envisagé son approche transfrontalière grâce à l'étude de feuilles de route organisationnelles, nous formulerons notre problématique du point de vue des sciences de l'information et de la communication.

En termes informationnels, il s'agit de comprendre comment s'organise la visualisation conjointe de données hétérogènes fournies par des systèmes nationaux de surveillance épidémiologique, l'objectif consistant à élaborer des systèmes communs d'information, d'alerte précoce ou de veille. En effet, si des efforts sont observés dans la mise à disposition d'informations partagées à l'échelle transfrontalière, ils se heurtent à de nombreux problèmes spécifiques (notamment pointés en France par la Mission opérationnelle transfrontalière) : obstacles liés aux cadres légaux et administratifs des pays concernés, méconnaissance des territoires transfrontaliers, approche politique marginale et souvent à court terme en matière de coopération transfrontalière, financements sur projets limités dans le temps et surtout manque d'informations et de données harmonisées. En termes communicationnels, il s'agit de comprendre comment se construit une vision partagée de situations épidémiologiques, l'objectif étant d'élaborer une identité éditoriale des plateformes à l'issue d'un travail interdisciplinaire.

Un corpus de 10 plateformes dédiées à la surveillance épidémiologique (en Europe, Amérique latine, Afrique et Antilles) fait l'objet d'une analyse discursive et sémiotique. L'observation est menée de manière à confronter le corpus aux prescriptions d'ouverture et de partage des données à l'échelle transfrontalière et à déceler les efforts et limites d'ouverture des données transfrontalières.

Le cadre théorique et méthodologique mobilisé fait appel à la classification des archives de l'histoire immédiate (Soulet, 2012), au concept d'archive redéfini en sémiotique appliquée aux médias comme un « fond textuel évolutif » (Stockinger, 2012) et à l'analyse sémio-pragmatique de portails d'*open data* (Paquienséguy, 2018).

Après dépouillement des contenus des plateformes, nous sommes en mesure de déterminer :

- les principaux apports du partage de données transfrontalières : valorisation d'organisations émergentes, identification de gisements de données épidémiologiques transfrontalières, typologie des projets et plateformes considérés, valorisation de compétences sur la comparabilité de données en contexte transfrontalier ;
- les principales limites concernant la visualisation conjointe des données à l'échelle transfrontalière : volonté manifeste de partage mais faible ouverture des données en matière de santé transfrontalière, qui fait écho aux pratiques prudentes en matière d'ouverture des données de santé (Guigue & Richard, 2014) ; manque (récurrent mais non systématique) d'explicitation des dispositifs mis en place (nature et source des données, saisie des données, mise à disposition et analyse des données, formulation de l'état d'avancement et des perspectives).

En complément, des entretiens sont menés avec des porteurs et acteurs de projets d'ouverture des données épidémiologiques à l'échelle transfrontalière. Ciblés sur les stratégies des promoteurs, les contextes d'usage, les flux des données et les tensions vécues, ils permettent d'alimenter les réflexions en sciences de l'information et de la communication sur les méthodes et pratiques de recherche sur les données géographiques et en sociologie des sciences sur le processus de mise en bases de données. En questionnant les arguments récurrents de justification de la mise en œuvre de ces plateformes, nous précisons ce que recouvre l'« égalité informationnelle des territoires » (Pinède et al., 2017) en contexte épidémiologique transfrontalier. En questionnant les difficultés rencontrées pendant l'élaboration de ces plateformes, nous identifions les « moments frictionnels » (Jaton & Vinck, 2016) caractéristiques de leur dynamique de construction.

Références bibliographiques

- Guigue L., Richard C., 2014, « Le *big data* en santé préfigure-t-il la "médecine 3.0" ? », *Hegel*, 3(3), p. 273-278 [doi.org/10.4267/2042/54093].
- Jaton F., Vinck D., 2016, « Processus frictionnels de mises en bases de données », *Revue d'anthropologie des connaissances*, 10(4), p. 489-504 [doi.org/10.3917/rac.033.0489].
- Paquiénéguy F., Dymytrova V., 2018, « *Open data* et métropoles, les enjeux d'une transformation à l'œuvre. Analyse sémio-pragmatique d'un corpus de portails métropolitains », *Questions de communication*, 34(2), p. 209-228 [doi.org/10.4000/questionsdecommunication.15818].
- Pinède N., Noucher M., Gourmelon F., 2017, « De l'*open data* à l'*open science* : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques », *Revue française des sciences de l'information et de la communication*, n° 11 [doi.org/10.4000/rfsic.3200].
- Soulet J.-F., 2012, *L'histoire immédiate. Historiographie, sources et méthodes*, Paris, Armand Colin.
- Stockinger P., 2012, *Analyse des contenus audiovisuels : métalangage et modèles de description*, Paris, Lavoisier.

L'autrice

Marie-Hélène HERMAND

Université Bordeaux Montaigne – EA4426 Médiations, informations, communication et arts (MICA)

marie-helene.hermand@u-bordeaux-montaigne.fr

Construction d'une source en *open data* sur le recours aux soins de première ligne

Jean-Marc MACÉ, Katerina KONONOVICH

Mots clés

diagnostic territorial – territoires vécus – accès aux soins – géographie de la santé – aménagement sanitaire du territoire

Le diagnostic territorial est à la territorialisation sanitaire ce que sont les fondations d'un immeuble en architecture. En effet, pour guider les politiques publiques dans leurs priorités d'accès aux soins, le diagnostic territorial est primordial. Il permet de rechercher la meilleure adéquation de l'offre de soins à la demande effective de soins de première ligne. Cet objectif de transformation de données de santé en accès restreint en *open data* sur le recours aux soins de première ligne s'inscrit dans cette volonté de conforter le diagnostic territorial et in fine la territorialisation sanitaire.

Méthodologie

Pour y parvenir, les chercheurs se sont appuyés sur les bases de données provenant du programme de médicalisation des systèmes d'information (PMSI), de la statistique annuelle des établissements (SAE) et du système national d'information inter-régimes de l'Assurance maladie (SNIIRAM).

Si les deux premières bases de données (SAE et PMSI) restent sujettes à caution quant au recensement de l'activité des services des urgences, la base du SNIIRAM apparaît plus exhaustive puisqu'elle collige l'ensemble des forfaits d'accueil et traitement des urgences (ATU) remboursés pour chaque usager. Toutefois, une partie de l'activité des services des urgences reste inconnue comme par exemple les patients non remboursés par l'Assurance maladie (étrangers, personnes hors protection universelle maladie).

Face à ces contraintes dans le rapprochement des trois bases de données, l'arbitrage choisi pour conduire cette recherche s'appuie donc sur un dénombrement de l'activité des services des urgences selon trois phases :

- d'abord les ATU sont identifiés depuis la base du SNIIRAM ;
- puis les hospitalisations issues du service des urgences de chaque établissement sont captées par l'intermédiaire de la base SAE ;
- enfin, la somme de ces deux activités constitue le total de références de l'activité de ces différents services d'accueil des urgences en France métropolitaine.

Les territoires de référence retenus sont les territoires vécus des services des urgences hospitalières sur la France métropolitaine établis depuis la technique des « flux relatifs » pour mettre en place une typologie reposant sur des analyses multivariées (analyse en composantes principales). L'ensemble de ces mesures ont permis de construire l'*open*

data sur le recours aux urgences de première ligne dont l'accès aux données au grand public est disponible gratuitement en ligne¹.

La disponibilité des données au grand public

Les données concernant les flux vers les services des urgences sont visualisables sous forme d'infobulle au niveau de chaque commune. L'utilisateur peut ainsi parcourir la carte de France et choisir les informations sur les 35 500 communes. Cette interface cartographique permet également d'obtenir au niveau des territoires vécus des services des urgences des données sanitaires agrégées comme la densité de médecins généralistes, le nombre de consultations chez les médecins généralistes, l'activité moyenne de consultations par médecin généraliste, la consommation moyenne de consultations par médecin généraliste, le nombre de passages aux services des urgences (avec hospitalisation ou non) et l'accès en minutes aux services des urgences depuis chaque commune par un jeu interactif des échelles géographiques allant du niveau national au niveau infra communal.

Les difficultés rencontrées pour la mise à disposition en *open data*

La transformation de données de santé en accès restreint vers une source *open data* font l'objet de grandes difficultés.

Rendre disponible les données traitées au grand public passe par deux grandes étapes :

- L'autorisation des données par le Health Data Hub² et la CNIL, qui reste un vrai parcours du combattant : 2 à 3 ans d'attente entre le dépôt du dossier et la mise à disposition des données (hors période sanitaire actuelle) ;
- la maintenance du site internet et son maintien en ligne reste très onéreux (5 000 €/an) et nécessite une recherche permanente de fond.

Les auteurs

Jean-Marc MACÉ

CNAM – EA4603 Laboratoire interdisciplinaire de recherche en sciences de l'action (Lirsa)
macejmarc@gmail.com

Katerina KONONOVICH

HESAM – EA4603 Laboratoire interdisciplinaire de recherche en sciences de l'action (Lirsa)
katerina.kononovich@gmail.com

1 fr.geoconcept.com/les-urgences-en-france-metropolitaine-etude-sur-la-geographie-de-la-sante

2 Plateforme des données de santé, le HBH est une infrastructure officiellement créée le 30 novembre 2019, destinée à faciliter le partage des données de santé issues de sources très variées afin de favoriser la recherche.

Table des matières

Appel à communications – Accès et utilisation des données en libre accès en géographie de la population, de la santé et des mobilités.....	1
Programme	3
Session 1. Géographie de la mobilité et de la population	5
Julie VALLÉE – Diffuser et croiser des données libres en géographie : un retour d’expérience autour de l’outil <i>Mobiliscope</i>	7
Boris MERICKSKAY, Florent DEMORAES – Les données DVF en <i>open data</i> comme matériel de recherche pour les géographes : propositions méthodologiques à travers l’exemple de la Bretagne et la métropole rennaise	9
Marianne GUÉROIS, Timothée GIRAUD, Ronan YSEBAERT – Apports et limites de la base <i>OpenStreetMap</i> pour l’analyse de l’accessibilité territoriale et de la localisation d’équipements	11
Nicolas OVTRACHT, Camille SIEPER, Boris MERICKSKAY, Florent DEMORAES, Sarah DUCHÉ, Kamila TABAKA – Construction d’une nomenclature unifiée des aménagements cyclables en France et en Colombie à partir de données institutionnelles ou en libre accès : enjeux et perspectives	13
Patrice TISSANDIER, Jérémy BÉTHUNE – Utilisation de données en libre accès visant à la création d’un outil d’aide à la décision pour l’implantation d’espaces d’enseignement à distance	15
Session 2. Géographie de la santé	17
Maxime THORIGNY, Leila BOUAZZI, Alexandre DENOYER, Stéphane SANCHEZ – Étude exposition aux pesticides et maladies neurodégénératives.....	19
Lara-Maria WAKIM, Florent OCCELLI, Damien CUNY, Caroline LANIER, Francis DOUAY, Annabelle DERAM – Maladie de Crohn et pollution des sols	21

Delphine BROUSMICHE, Caroline LANIER, Annabelle DERAM, Florent OCCELLI, Michaël GENIN, Aline MEIRHAEGHE, Damien CUNY – Développement d’indices composites spatialisés pour l’étude du lien entre environnement au sens élargi et survenue de pathologies chroniques : apports et limites de l’ <i>open data</i>	23
Martin PAUMELLE, Florent OCCELLI, Caroline LANIER, Damien CUNY, Annabelle DERAM – Utilisation de données en libre accès pour caractériser le lien entre insuffisance rénale chronique terminale et environnement	25
Marie-Hélène HERMAND – Les <i>open data</i> territorialisées à l’échelle transfrontalière : prescriptions, enjeux et réalités à l’aune d’un corpus de plateformes de surveillance épidémiologique	27
Jean-Marc MACÉ, Katerina KONONOVICH – Construction d’une source en <i>open data</i> sur le recours aux soins de première ligne.....	29
Table des matières	31