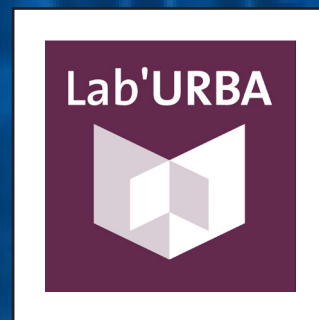


Atelier méthodologique 2026
Lab'URBA
4^{ème} Édition



LA FABRIQUE DES DONNÉES DANS ET POUR LA RECHERCHE

20 mars 2026
12h-17h

Cité Descartes
Bâtiment Bienvenüe
14-20, boulevard Newton
77420 Champs-sur-Marne

Organisation : Aymeric Bouchereau (UPEC), Juliette Morel
(UPEC), Léa Prost-Lançon (UPEC), Francesca Artioli (UPEC)

Contact : lea.prost@u-pec.fr



PRÉSENTATION DE LA THÉMATIQUE 2026

La quatrième édition de l'atelier méthodologique du Lab'URBA est dédiée au travail des données dans la recherche et en particulier dans les études urbaines. Loin d'être un déjà là acquis (malgré ce que l'étymologie du mot « données » peut laisser croire), la donnée est fabriquée à partir de traitements concrets. Cet atelier méthodologique sera le moment d'interroger les formes de bricolages qui s'opèrent pour fabriquer les données de nos travaux de recherche.

Les données sont la plus petite partie à partir de laquelle sont produites les connaissances. L'OCDE en donne une définition objectiviste, comme des « enregistrements factuels (chiffres, textes, images et sons) qui sont utilisés comme sources principales pour la recherche scientifique » (OCDE, 2007). Au contraire de cette définition positiviste, et en même temps que se développe une « euphorie médiatique » autour des big data dans les années 2010, les *Critical data studies* établissent qu'« en pratique, les données ne sont pas de simples preuves de phénomènes, elles sont des phénomènes en soi. Les “big data” ne sont jamais “brutes” » (Dalton & Thatcher, 2014).

Dans tous les cas, qu'elles soient fabriquées par le·a chercheur·e ou récupérées à partir d'une autre source, les données occupent un volet à part entière de la recherche en sciences humaines et sociales (Denis & Goëta, 2017). Elles peuvent être issues d'expérimentations, de résultats d'analyse par des instruments, de collectes sur le Web, d'observations, d'enquêtes, d'entretiens, de questionnaires, ou encore d'archives. Elles nécessitent un certain nombre de traitements : l'identification et l'exploration de potentielles données, l'extraction pour former un corpus, le nettoyage, le (re)-formatage et la structuration, la mise en intelligibilité, etc. Ces traitements sont parfois complexes, imprégnés de technique, faits d'arbitrages, marqués par certaines orientations voire des biais – qu'il est important de questionner tout au long du cycle des données dans la recherche en SHS, de leur fabrication à leur circulation, en passant par leur utilisation pour produire de la connaissance. Le tournant numérique a en outre contribué à multiplier les modes d'obtention, les formats et les types de données ainsi que la taille des corpus de données, numérisées ou nativement numériques (Sauret, 2022). La circulation des données de la recherche tend enfin à se massifier, sous l'effet notamment des politiques d'incitation à la science ouverte et les potentiels de diffusion en ligne. Cela implique qu'il faut intégrer dans nos méthodologies de recherche le fait de rendre les données produites et/ou utilisées accessibles, interopérables et réutilisables (Comité pour la science ouverte, 2024 ; principes FAIRE). C'est tout cela que le quatrième atelier méthodologique du Lab'URBA entend explorer.

Comité pour la science ouverte. (2024). Science ouverte—Données de la recherche (Passeport pour la science ouverte, p. 40). Ministère de l'Enseignement supérieur et de la Recherche. <https://www.ouvri.lascience.fr/wp-content/uploads/2024/03/24-02-28-Donnees-FR-IMPR.pdf>

Dalton, C.M. et Thatcher, J. (2014) « What Does A Critical Data Studies Look Like, And Why Do We Care? », *Society and space* [Preprint]. Disponible sur: <https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care> (Consulté le: 20 avril 2023).

Denis, J., & Goëta, S. (2017). La fabrique des données brutes : Le travail en coulisses de l'open data. In C. Mabi, J.-C. Plantin, & L. Monnoyer-Smith (Éds.), *Ouvrir, partager, réutiliser : Regards critiques sur les données numériques*. Éditions de la Maison des sciences de l'homme. <https://doi.org/10.4000/books.editionsmsmh.9050>

Sauret, N. (2022). Intelligence artificielle & Sciences humaines et sociales (SHS) : Opportunités, défis et perspectives. *I2D - Information, données & documents*, 1(1), 97-103. <https://doi.org/10.3917/i2d.221.0097>

OCDE. (2007). Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics (p. 53). Éditions OCDE. <https://doi.org/10.1787/9789264034020-en-fr>



PROGRAMME

- 12h** Accueil / Déjeuner collectif
- 13h30** « Le capital de logement », Arnaud SIMON (UPEC)
- 14h15** « De la méthode à l'objet : approche réflexive d'une recherche-action sur la gouvernance territoriale de la donnée », Marie-Hélène GAUTHIER (UPEC)
- 15h** Pause
- 15h15** « Entre observations et questionnaires : extraire, structurer et mettre en relation des jeux de données de sciences participatives dans une recherche interdisciplinaire », Camille TOURTELIER (Univ. Paris Saclay, Ville de Paris)
- 16h** Discussion collective sur la place des données dans nos recherches et dans le laboratoire

RÉSUMÉS

« Le capital de logement », Arnaud SIMON

En s'appuyant sur la base des transactions de logements DV3F et sur les Fichiers Fonciers ce travail estime la valeur du capital de logements en France et il en analyse les distributions. La concentration spatiale du capital est étudiée au niveau départemental et communal, la distribution institutionnelle l'est entre les particuliers, le parc social, les SCI/Investisseurs et les collectivités locales. L'estimation du flux annuel de la richesse immobilière héritée et les enjeux fiscaux sont aussi discutés. Enfin, une attention particulière est portée à la distribution du capital entre les classes d'âges, en cherchant à qualifier spatialement, démographiquement et économiquement les situations de surplus, c'est-à-dire de sur-captation capitaliste.

« De la méthode à l'objet : approche réflexive d'une recherche-action sur la gouvernance territoriale de la donnée », Marie-Hélène GAUTHIER

Cette communication portera sur les enjeux et les effets de la collecte et du traitement des données, lorsque la donnée constitue elle-même l'objet de recherche. Les résultats proviennent d'une thèse CIFRE, réalisée à la Région Normandie, portant sur la coordination d'acteurs territoriaux (collectivités territoriales, entreprises, acteurs académiques) pour valoriser la donnée à l'échelle régionale. A travers une approche réflexive, les résultats mettront en évidence un effet miroir entre l'application de la méthodologie et ce qu'elle permet de comprendre de la gouvernance territoriale de la donnée, tout en y participant.



« Entre observations et questionnaires : extraire, structurer et mettre en relation des jeux de données de sciences participatives dans une recherche interdisciplinaire », Camille TOURTELIER

La présentation portera sur la pratique du nourrissage des oiseaux en milieu urbain, à travers un programme de sciences participatives appelé BirdLab , où des citoyen.ne.s reportent leurs observations d'oiseaux à la mangeoire. L'approche est interdisciplinaire, combinant écologie & sciences humaines et sociales, comme en témoignent les jeux de données mobilisés (identification, exploration et extraction). Le premier jeu de données, de nature écologique et pré existant à ma thèse, contient plus de 11 ans de données d'observations d'oiseaux aux mangeoires grâce à BirdLab, nécessitant un certain nombre de traitements pour les rendre intelligibles. Le deuxième jeu de données, porté lui sur l'aspect humain, a été construit et récolté via un questionnaire diffusé pendant la thèse.

Une fois les traitements opérés sur chacun des jeux de données respectifs, l'objectif est de les mettre en relation (structuration), afin de mutualiser les enseignements de chacun et permettre une compréhension globale et socio-écologique de la pratique du nourrissage des oiseaux et des interactions oiseaux-humains. Le projet de recherche s'appuyant sur des données de sciences et recherches participatives et étant financé dans un cadre d'une thèse CIFRE, la diffusion des résultats et l'importance du retour sur investissement (circulation et ouverture) sont des enjeux particulièrement réfléchis.

Il s'agira de discuter de ces différentes étapes sur la fabrique et l'usage des données de ma thèse, en appuyant sur la spécificité de l'ancrage du projet à la fois interdisciplinaire, mobilisant des données de sciences participatives, et s'inscrivant dans un financement CIFRE par la Ville de Paris.



