

APPEL À PROJETS 2024 - DONNÉES & OPEN DATA

FÉDÉRATION DE RECHERCHE AGORANTIC «CULTURE, PATRIMOINES, SOCIÉTÉS NUMÉRIQUES »

Attention :

- **Ne pas dépasser 5 pages**
- Transmettre le fichier au format PDF intitulé : ACRONYME-AAP-Données-Agorantic-2023
- Envoyer le fichier à fr-agorantic@univ-avignon.fr **avant le 24 novembre 2023.**

Titre	Interoperability of Land registry and Entreprise Data
Acronyme	InterLEnD
Nom du/des porteur(s)	Antoine Peris Pierre-Henri Morand
Coordonnées du/de la gestionnaire de laboratoire	Alexia Bouffel (alexia.bouffel@univ-avignon.fr)
Laboratoires associés	ESPACE - JPEG
Budget demandé	4620 €
Résumé <i>Max. 1 000 caractères espaces compris</i>	Le projet InterLEnD a pour objectif de collecter, structurer et analyser les données sur la gouvernance et les bénéficiaires des Sociétés Civiles Immobilières (SCI) possédant des logements en France. L'analyse de ces données disponibles en Open Data permettra l'identification de propriétaires réels dont les patrimoines sont actuellement invisibilisés par la structure des données foncières. Par le croisement de différentes bases de données récemment ouvertes par l'administration publique, ce projet permettra de nourrir plusieurs actions de recherche engagées actuellement au sein de l'université : d'une part les travaux sur les structures locales de propriété développés à l'UMR ESPACE, en permettant une analyse plus fine des propriétaires de logements ; d'autre part les travaux sur les données de la commande publique (et notamment base FOPPA développée dans le cadre de l'ANR Détecter la Corruption dans les Marchés Publics), en enrichissant cette dernière des liens interpersonnels qui se nouent à travers les SCI.

1. Contexte, positionnement, objectif(s)

Une société civile immobilière (SCI) est une forme juridique permettant d'acquérir et gérer un patrimoine à plusieurs. Si la part du parc de logement détenue par ce type de structure est relativement faible, une proportion significative des logements possédés par des SCI se trouvent dans des marchés extrêmement valorisés tels que les centres des grandes métropoles ou les zones littorales. Par ailleurs, depuis une dizaine d'années, la détention de logements via ce type de véhicule progresse fortement au détriment de la propriété occupante et de la multipropriété de plein nom. Cette dynamique concerne les marchés tendus comme détendus.

Cependant, cette forme juridique identifiable comme telle dans les bases de données foncières recoupe une multitude de réalités et de propriétaires réels. En effet, les SCI peuvent servir :

- de véhicules pour des gestionnaires d'actifs agissant pour le compte d'investisseurs institutionnels ;
- à des montages pour des multipropriétaires qui ont professionnalisé leur activité ;
- à des montages pour des petits propriétaires (dans le cadre d'investissements locatifs, pour faciliter des transmissions intergénérationnelles ou répartir la propriété entre plusieurs ménages) ;
- à des montages permettant la détention indirecte d'immobilier via des sociétés off-shore (dans le but de se soustraire à l'impôt sur la fortune immobilière, de blanchir des capitaux mal-acquis ou de contourner des sanctions internationales).

Raffiner la catégorie des SCI dans les données foncières constitue donc un enjeu important pour améliorer la connaissance des dynamiques actuelles des marchés du logement qui font l'objet de tensions croissantes dans un contexte de crise d'abordabilité.

Depuis 2016, la loi Sapin II sur la transparence, la lutte contre la corruption et la modernisation de la vie économique impose la tenue d'un registre des bénéficiaires effectifs (RBE). Depuis janvier 2023, les bénéficiaires des sociétés commerciales et civiles doivent être déclarés auprès de l'INPI. En exploitant les données issues de ces enregistrements ainsi que la gouvernance des SCI, il est ainsi possible de les rattacher à des propriétaires réels. Ce croisement d'information permettrait de rattacher les SCI à des propriétaires physiques quand il s'agit de SCI familiales ou de montages permettant la détention indirecte d'immobilier. Il permettrait également de reconstituer des portefeuilles d'actifs immobiliers à l'échelle de groupes dans les cas de foncières, de gestionnaires d'actifs ou de multipropriétaires ayant professionnalisé leur activité.

En croisant différentes bases sur la propriété foncière et la structuration des entreprises accessibles en Open Data via le Registre national des entreprises (RNE) il est ainsi possible de lever un verrou d'opacité actuellement présent dans les données foncières et d'améliorer la transparence des marchés du logement.

Outre la transparence qu'elle permettrait d'atteindre sur les données foncières, l'identification des personnes physiques bénéficiaires effectifs et dirigeant de ces structures offre de surcroît une opportunité non encore exploitée de mettre à jour les relations interpersonnelles qui s'y décrivent implicitement. Liens familiaux, liens maritaux et liens d'affaires se révèlent ainsi à travers ces structures. En identifiant par ce biais l'existence et l'intensité de ces liens, ces données ont vocation à améliorer les données relationnelles (entre structures attributrices de marchés publics et entreprises attributaires) qui sont étudiées dans le cadre de l'ANR DeCoMaP (portée à Pierre-Henri Morand, et dont les laboratoires JPEG et LIA sont partenaires).

Le projet mobilisera :

Antoine Peris, MCF en géographie, UMR ESPACE, Avignon Université

Pierre-Henri Morand, PR en économie, JPEG, Avignon Université

Laure Casanova Enault, MCF en géographie, UMR ESPACE, Avignon Université

Kévin Poisson, doctorant en géographie, UMR Géographie-cités, Université de Paris

2. Questionnement scientifique en rapport avec l'intitulé de l'appel

Le projet InterLEnD s'insère dans l'axe 3 de la FR Agorantic « Politique(s), transparence et éthique », notamment dans sa volonté de développer des outils automatiques pour recouper et exploiter des sources de données existantes.

Depuis mars 2021, la Direction Générale des Finances Publiques a ouvert des données sur les locaux et les parcelles détenues par des personnes morales (organisations publiques et sociétés publiques et privées). Le contenu de ces cette base de données est relativement sommaire et difficilement compréhensible par le grand public mais il présente un fort potentiel d'interconnexion avec d'autres sources de données ouvertes. Tout d'abord, par l'interconnexion avec le cadastre (PCI), il est possible de localiser ces propriétés. Ensuite, l'interconnexion avec les données publiques sur les entreprises

(RNE) rend possible la reconstitution de montages patrimoniaux complexes et des liens entre les personnes partageant des parts ou des responsabilités dans des SCI.

Cette mise à disposition de données résulte de plusieurs facteurs, notamment la numérisation des administrations, la réponse aux injonctions de transparence émanant du grand public comme des institutions européennes, et la volonté de favoriser l'émergence d'un écosystème d'entreprises exploitant les données foncières. De manière collatérale, il ouvre de nouvelles perspectives aux chercheuses et chercheurs en sciences sociales qui travaillent sur le fonctionnement des marchés fonciers-immobiliers, les structures locales de propriété et le fonctionnement des marchés publics.

Les données que nous exploiterons dans ce projet présentent cependant certaines limites (pas d'obligation d'enregistrer des bénéficiaires détenant moins de 25% des parts des entreprises, données manquantes, etc.). Leur exploitation permettra ainsi d'évaluer dans quelle mesure ces règles et manquements contribuent au maintien d'une certaine opacité des marchés fonciers-immobiliers.

3. Méthodologie

La récupération des données sur les bénéficiaires effectifs et la gouvernance des entreprises peut-être effectuée en requêtant l'API Entreprises mise à disposition par l'Institut National de la Propriété Intellectuelle (INPI). Cette API permet d'obtenir, pour une date donnée, des informations sur le conseil d'administration, les gestionnaires et les bénéficiaires des sociétés enregistrées en France. Un enjeu méthodologique est donc d'effectuer un nombre important de requêtes tout en respectant les limites imposées par le service, et donc de distribuer et d'étaler le processus de collecte de données dans le temps.

Dans un second temps, il sera nécessaire de concevoir une méthode permettant de gérer la variabilité dans les enregistrements des noms des gestionnaires et des bénéficiaires des SCI afin d'identifier des personnes uniques. Ce travail devra convoquer à la fois des méthodes d'analyse de chaînes de caractères et un recours à de l'analyse des réseaux formés par les liens entre les personnes et les entreprises.

Enfin, un travail de spatialisation des résultats (via l'interconnexion aux données cadastrales ouvertes) et aux données publiques sur les transactions foncières devra être effectué pour renseigner les dynamiques en cours au sein des marchés locaux.

4. Résultats attendus et caractère innovant de la recherche

- Interconnexion de trois bases de données : le RNE, le Fichier des locaux des personnes morales et la base FOPPA.
- Publication d'une nouvelle base de données sur les différents types de SCI possédant des logements à l'échelle nationale, leur inscription dans des structures patrimoniales complexes, leurs gestionnaires et bénéficiaires.
- Publication d'un data paper sur l'interconnexion des bases de données.

5. Dimension interdisciplinaire

Le projet InterLEnD est interdisciplinaire dans la mesure où il nécessite de croiser l'expertise de géographes et d'économistes dans un cadre relevant des sciences sociales computationnelles.

- D'abord, la collecte des données devra reposer sur un algorithme performant et distribué qui allie vitesse d'exécution et conformité aux règles énoncées par l'organisme maintenant le RNE.
- L'identification unique des acteurs des SCI reposera elle aussi sur des méthodes innovantes d'appariement flou et d'analyse de réseaux formés par les sociétés,

leurs bénéficiaires et leurs gestionnaires.

- L'expertise économique sera convoquée pour la compréhension des structures des entreprises et des relations de propriété.
- Enfin, l'expertise de géographes permettra la spatialisation des données et leur exploitation pour une meilleure compréhension des marchés fonciers-immobiliers locaux.

6. Partenariats extérieurs envisagés

- Kévin Poisson, doctorant contractuel à l'UMR Géographies-cités, travaillant sur les entreprises détenant de l'immobilier d'activité
- Antoine Paccoud et Laure Casanova Enault, Projet ANR DISTRILAND (Encastrement socio-juridique des marchés fonciers : Approche comparée de la structure de propriété et des systèmes de régulation) - Luxembourg Institute of Socio-Economic Research & Avignon Université

7. Valorisation (si prévue)

Publication d'une base en accès libre qui pourra par ailleurs être intégrée aux données de la Chaire GIF et aux données issues du projet BORIS.

8. Expression des besoins en assistance informatique

Expertise pour solution technique à la collecte massive des données compte-tenu de contraintes quantitatives liées à l'API mobilisée dans le cadre du projet
Interfaçage des données avec les bases BREF et FOPPA

9. Budget (€) prévisionnel *

	Brève description	Montant
Missions	Déplacements stagiaire et Antoine Paris	600
Consommables, petits matériels**	6 x Raspberry pi 4GB 6 x carte SD 32 GB 6 x alimentations 12v (nécessaires pour constituer des serveurs de collecte des données)	540
Stages***, vacations		3480
Budget total		4620
Budget demandé à Agorantic		4620

* Veuillez modifier les catégories de dépenses si besoin - ajoutez/supprimez des lignes à votre convenance

** Petit matériel ne dépassant pas les 600€

***Gratification de stage obligatoire au-delà de 2 mois - prévoir environ 580€ par mois

NB : si le projet a déjà fait l'objet d'un financement lors d'un précédent AAP : justifier la nouvelle demande et présenter les évolutions du projet.