

APPEL A PROJETS RECHERCHE 2021

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

Attention :

- Ne pas dépasser 5 pages
- Transmettre le fichier au format PDF intitulé : ACRONYME-AAP-Agorantic-2021.pdf
- Envoyer le fichier à agorantic@univ-avignon.fr avant le vendredi 16 novembre 2020.

Titre	Le Mythe d'un marché Immobilier Transparent grâce à l'Open data
Acronyme	MITO
Nom du/des porteur(s)	Guilhem Boulay
Laboratoires associés	UMR 7300 ESPACE (L. Casanova Enault, D. Josselin, P. Le Brun), EA 2151 LMA (D. Blanke), EA 4128 LIA (R. Dufour)
Budget demandé	5520 €
Résumé Max. 1 000 caractères espaces compris	<p>Début 2019, la DGFIP (Ministère de l'Economie, des Finances et de la Relance) a ouvert l'accès à la base DVF (Demande de Valeurs Foncières) qui recense l'intégralité des transactions foncières et immobilières de France, qualifiées par plusieurs variables (date, montant et régime juridique de la transaction, caractéristiques et localisation du bien, etc.).</p> <p>Dans un contexte de tension durable des marchés immobiliers, cette ouverture des données a été saluée par le gouvernement comme un moyen d'accroître la transparence des marchés et d'en tempérer les prix.</p> <p>Les deux hypothèses de MITO remettent en cause ce discours : l'ouverture des données aurait surtout alimenté la croissance de nouveaux secteurs économiques (notamment les start-ups de la <i>proptech</i>) et la difficulté de manipulation de DVF rendrait les utilisateurs captifs des traitements méthodologiquement opaques effectués par la <i>proptech</i>.</p> <p>Ces hypothèses seront testées par des méthodes mixtes et pluridisciplinaires : exploration des données Twitter et des réseaux qu'elles signalent, entretiens auprès des usagers de l'open DVF et analyse des biais statistiques liés à l'agrégation des données.</p>

1. Contexte, positionnement, objectif(s)/problématique et cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques »

Le projet MITO traite d'objets et repose sur des méthodologies au cœur des préoccupations de la FR Agorantic.

L'open data et la multiplication des informations sur Internet, ainsi que le recours massif à Internet pour rechercher des informations est un des marqueurs des « sociétés numériques ». Cela est particulièrement vrai dans le domaine de l'immobilier : tous les mois, 29 % des utilisateurs individuels français d'Internet

consultent des sites immobiliers (avec pour motif dans 87 % des cas la recherche d'informations sur les prix) (Lefebvre, 2015).

Par ailleurs, le protocole méthodologique qu'entend suivre MITO repose sur des méthodes mixtes (entretiens, analyse de données en ligne, traitements statistiques) et sur plusieurs disciplines (géographie, statistique, informatique).

Il relève à ces titres des axes 1, 3 et 5 de la FR.

Ce projet fait en partie suite au travail effectué dans le cadre d'un projet Agorantic obtenu lors de l'AAP 2014 (projet « Observer les marchés par l'Internet : représentation et qualité de l'information statistique sur les marchés fonciers-immobiliers en France et en Espagne ») et notamment valorisé par une publication collective internationale : Boulay G., Blanke D., Casanova Enault L., Granié A., 2020, « Moving from market opacity to methodological opacity: Are web data good enough for French property market monitoring ? », *The Professional Geographer*, DOI : 10.1080/00330124.2020.1824678

2. Questionnement scientifique

Depuis les années 1960-1970, l'économie standard insiste énormément sur le rôle de l'information dans le fonctionnement des marchés. Une meilleure information permettrait d'accroître la transparence des marchés et d'améliorer la fixation des prix et l'allocation des ressources. Ce type de position s'est fortement développé dans le contexte immobilier, comme en témoigne le livre sur la crise des *subprime* de R. J. Shiller. Le « Nobel d'économie » y affirme notamment que « *[p]romoting an improved information infrastructure – the knowledge base used by people and firms to manage their financial affairs – is not only good policy but also good economics* » (Shiller, 2011). Ces positions standard, normatives, ont fortement percolé dans les politiques publiques sous l'effet de la néo-libéralisation de l'Etat (Dardot, Laval, 2009), comme en témoigne l'ouverture des données DVF. Cela ne va pas toutefois sans poser deux questions.

D'une part, quelle que soit l'opinion qu'on ait sur la justesse des positions théoriques ou quelle que soit la sincérité de l'Etat lors de l'ouverture des données, on sait que toute réforme politique peut s'interpréter comme un nouveau champ de contraintes et de ressources. Se pose donc la question de savoir qui s'est emparé des données nouvellement en ligne, et pour quoi faire (Kitchin, Lauriault, 2014). Des travaux antérieurs ont ainsi montré l'importance de l'open data dans de nombreuses stratégies économiques, de start-ups comme d'acteurs économiques plus établis (Boulay *et al.*, 2020).

D'autre part, quelle que soit la qualité des traitements, la possibilité d'observer de manière neutre et immédiate les marchés pose question. En effet, comme beaucoup d'autres données socio-économiques, les données sur les marchés immobiliers relèvent plus de « conventions d'équivalence » que de « mesures » (Desrosières, 2001). A ce problème de fond s'ajoutent de multiples sources de biais purement méthodologiques : choix de valeurs centrales (qui présupposent des distributions statistiques), choix des mailles spatiales sur lesquelles calculer des indicateurs (qui négligent les problèmes d'agrégation et de zonage mis en évidence par le MAUP), choix enfin de segmentation des marchés (qui négligent le caractère construit et nécessairement arbitraire des données) (Josselin, 2019, Casanova Enault *et al.*, 2019).

3. Méthodologie et résultats attendus

Ces deux questions scientifiques seront abordées par quatre types de méthodes :

- **Traitement automatique du langage pour l'exploration et l'analyse de données échangées dans le réseau social Twitter** : on collectera des données par mots-clés identifiés a priori avant d'utiliser un outil d'analyse automatique de polarité (positive / négatif / neutre) des tweets. Nous compléterons ce travail par l'extraction automatique des thématiques (transparence, empowerment citoyen, gisement de croissance, innovation technologique, etc.) développées dans les messages (soit par clustering, soit de façon supervisée en fonction des données annotées disponibles). **Résultats attendus** : connaissance des opinions relatives à l'ouverture de DVF et de leurs motifs.
- **Analyse des réseaux sociaux et professionnels révélés par les tweets** : on travaillera notamment à partir des relais de messages (retweets) en analysant les comptes utilisateurs liés au compte émetteur (followers/friends). L'analyse portera également sur le réseau complet des messages échangés. **Résultats attendus** : mise en évidence des polarités au sein des réseaux et des modalités de diffusion de l'information, notamment pour savoir estimer la porosité entre public et privé.
- **Campagne d'entretiens semi-directifs auprès des utilisateurs de l'open DVF** : on travaillera sur toute la chaîne de fabrication de l'information, depuis les producteurs de la donnée initiale (DGFIP) jusqu'aux utilisateurs finaux (ménages, professionnels de l'immobilier, collectivités territoriales et acteurs publics et parapublics), en passant par tous les intermédiaires qui retraitent et formatent la donnée et ses représentations (secteur de la *property technology* notamment). Concernant les intermédiaires, on étudiera également leurs méthodes et leurs outils (logiciels utilisés, choix des bases de données et des routines méthodologiques notamment). **Résultats attendus** : meilleure compréhension de la réintermédiation du marché immobilier par les producteurs de données et connaissance des modalités de ce *data manufacturing* (Sadowski, 2019).
- **Traitements statistiques et géostatistiques** : on travaillera sur les données DVF d'origine afin de mettre en évidence les biais induits par la spécificité des distributions statistiques de prix immobiliers (dissymétrie, queues de distributions et valeurs extrêmes), la distribution spatiale et temporelles des ventes (fortes disparités dans la couverture) et l'influence du MAUP (*Modifiable Areal Unit Problem* ou influence du zonage et de l'échelle retenus pour le calcul des paramètres statistiques : Josselin, 2019). **Résultats attendus** : mise en évidence des écarts de mesure et de la sensibilité des résultats aux choix statistiques, et proposition méthodologique pour améliorer la robustesse des estimations locales de prix immobiliers.

La réalisation de ces travaux est assurée par l'expertise thématique et méthodologique des membres du projet :

- **D. Blanke** : statisticienne, travaillant notamment en statistique des processus et en estimation robuste
- **G. Boulay** : géographe, spécialiste des marchés immobiliers et d'analyse spatiale
- **L. Casanova Enault**, géographe, spécialiste des marchés fonciers et d'analyse spatiale
- **R. Dufour**, informaticien, spécialiste de l'extraction d'information et de l'analyse des données Twitter
- **D. Josselin** : géographe, un temps rattaché à un laboratoire d'informatique, spécialiste de statistique spatiale et de robustesse statistique
- **P. Le Brun** : géographe (sujet de thèse : *Des cartes pour guider les capitaux : critique méthodologique et analyse de l'influence des modèles de segmentation des marchés immobiliers sur la production de l'espace urbain*)

4. Résultats attendus et caractère innovant de la recherche

On attend de ce projet :

- **une meilleure connaissance de la chaîne de production, de retraitement et de réutilisation de l'information géographique.** Ces résultats sont cruciaux dans la mesure où ils sont susceptibles d'éclairer la recherche sur l'influence des indicateurs sur des trajectoires économiques et spatiales de poids (performativité des indicateurs)
- **la proposition de méthodes robustes de traitement des prix immobiliers.** Ces résultats sont importants dans la mesure où la majorité des méthodes de traitement, y compris en économétrie, ignorent le MAUP et ses effets, et se concentrent seulement sur l'intégration massive de données. De manière plus fondamentale, les données immobilières présentent en outre des difficultés particulières, ce qui contribuera à l'amélioration des protocoles de traitement des données spatiales en général.

Le projet MITO est innovant dans la mesure où la littérature existante sur le sujet ou sur des sujets proches (données de mobilité par exemple) sépare systématiquement les approches qualitatives (rapport de force sur les marchés, privatisation de la donnée, etc.) et les approches quantitatives (qualité de la donnée, méthode de traitements, etc.). Or il est essentiel de montrer comment ces deux aspects interagissent si l'on souhaite progresser dans la compréhension de la performativité des indicateurs.

5. Dimension interdisciplinaire (champs disciplinaires associés) et cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques »

Le projet MITO réunit des collègues de différentes disciplines qui ont tou(te)s développé des compétences de travail interdisciplinaire, notamment dans le cadre des échanges au sein d'Agorantic. Certains ont déjà publié ensemble et continuent à le faire.

Ses objets répondent en outre complètement aux préoccupations de la thématique « Culture, Patrimoines, Sociétés Numériques » : open data, réintermédiation des marchés par les start-ups de l'Internet, information via les réseaux sociaux et les applications.

6. Partenariats extérieurs envisagés

Le projet MITO pourra s'appuyer d'une part sur une **très grande connaissance des réseaux professionnels de l'immobilier**, développée notamment grâce aux ateliers chercheurs-professionnels du RJCF (Réseau des Jeunes Chercheurs du Foncier, fondé notamment par G. Boulay et L. Casanova), aux journées d'études NUMIMMO (Numérique et Immobilier) financées par Agorantic en 2019 et 2020 (org. G. Boulay) et à la participation aux séminaires des grands réseaux publics du foncier et de l'immobilier (Rencontres des Acteurs Publics du foncier, Réseau des Opérateurs Fonciers, Assises Nationales du Foncier, etc.).

L'équipe dispose déjà de contacts ou de relations de travail stabilisées avec plusieurs entreprises, structures publiques et associations spécialisées : Métropoles de Lyon et d'Aix-Marseille-Provence, Etablissement Public Foncier PACA, bureaux d'études et consultants (notamment Adéquation, Claire Juillard et Alexandre Coulondre), start ups de la *proptech* à Paris et Lyon (Meilleurs Agents, e-attract, Vibris, Bien' Ici, etc.) ou associations (LIFTI, Fonciers en débat).

Il s'appuiera également sur les **réseaux scientifiques spécialisés** de chacun(e) de ses membres (GdR MAGIS par exemple).

Plusieurs canaux de valorisation sont envisagés : communications, publications internationales.

Le dépôt d'un **projet ANR ou international** (a priori avec des équipes issues des pays suivants : France, Espagne, Belgique, Pologne et Pays-Bas) sur la fabrication et la mobilisation des données immobilières est également envisagé à moyen terme (2022 a priori).

Budget (€)*		
	Brève description	Montant
Missions	<p>4 A/R sur deux jours à Paris pour une personne, et 1 autre à deux personnes (300 € / A/R)</p> <p>2 A/R de deux jours à Lyon pour une personne (250 € / A/R)</p> <p>2 A/R sur la journée à Marseille pour deux personnes (200 € / A/R)</p> <p>1 inscription au salon RENT (Real Estate & New Tech) Paris 2021 (200 €)</p>	<p>Paris : 1800</p> <p>Lyon : 500 €</p> <p>Marseille : 400 €</p> <p>RENT : 200€</p> <p>TOTAL MISSIONS : 2900 €</p>
Consommables / Documentation	Quelques livres anglophones sur le capitalisme à l'ère du digital, la statistique spatiale et la métrologie	300 €
Organisation de réunions	Intégralité de l'équipe en Avignon	Pas de frais
Stages**	4 mois de stage	2320 €
Budget total		5520 €
Cofinancements le cas échéant	On demandera à l'UMR ESPACE de subvenir aux frais de relecture d'un article en anglais par une traductrice spécialisée	Env. 500 €
Budget demandé à Agorantic		5520 €

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

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

NB : le texte projet devra mentionner si la thématique a été ou non déjà financée par un appel à projet Agorantic les années antérieures (sachant que cela n'est en aucun cas un motif de rejet du projet)