

INTENSITE DES FLUX ET DE LA DONNEE NUMERIQUE : UNE AUTRE APPROCHE DES MODES DE VIE METROPOLITAINS ?

Synthèse du séminaire 1 de l'axe 1 Modes de vie métropolitains et transitions

Territoire des plus connectés du fait de ses infrastructures et de son écosystème dense d'acteurs du numérique, la Métropole du Grand Paris (MGP) est un terrain d'observation privilégié de la transition numérique des villes. Les mobilités métropolitaines et les choix résidentiels permis par les nouvelles infrastructures de transport évoluent rapidement et complexifient la compréhension des modes de vie grand-parisiens. L'enjeu est alors d'interroger les modalités et les capacités de captage des données pour une meilleure compréhension de l'archipellisation des modes de vie. L'objectif est d'interroger l'épuisement des modes de saisie catégorielle et les indicateurs conventionnels pour la mesure de l'évolution des modes de vie métropolitains à l'aide de données numériques. Quels sont les sources et les distributeurs de ces données ? Qui les produit ? Quel serait l'intérêt à produire soi-même ces données ? Quel serait l'intérêt de redresser des données produites par d'autres ? Quelles précautions déontologiques prendre quant à la préservation des données personnelles ? Que signifient-elles pour les chercheurs et les agences d'urbanisme telles que l'Institut Paris Région ?

1 Capturer, produire et acheter des données massives

3,5 milliards de traces numériques
fournies par Leboncoin pour l'étude
menée sur l'Exode Urbain.

L'intégralité de notre environnement est « texturée » par le numérique, d'autant plus avec l'essor de l'IA appliquée aux méthodes d'apprentissage ou *deep learning*. L'enjeu affiché de l'utilisation des objets connectés et des données qu'ils produisent est alors bien d'améliorer et d'optimiser les usages et les modes de vie pour mieux paramétrer les pratiques des individus, comme en témoigne l'application #tousanticovid.

Dans quelles conditions, contextes et à quelles fins ces données sont-elles produites ? Comment y accéder ? Et comment pérenniser cet accès, surtout pour la recherche urbaine ?

Mytrafic est une entreprise qui propose de capturer et produire des données autour des mobilités, notamment piétonnes, en vue d'interroger les opportunités ou difficultés commerciales. La Métropole du Grand Paris en appelle à leurs services pour accompagner les communes, les aménageurs et les porteurs de projet commerciaux dans les choix stratégiques de localisation.

La technologie employée repose sur la géolocalisation par GPS, par exemple avec l'usage des téléphones portables, notamment grâce à des applications comme Google Maps, ou encore avec des traceurs GPS dédiés. Toutefois, les données produites témoignent de biais importants qui nécessitent de retravailler constamment la donnée.

La question se pose alors du financement et de l'achat de ce type de données, en plus du contexte dans lequel elles sont produites. Des partenaires de l'Institut Paris Région ont été convaincus par ce nouveau type d'enquête par GPS. Néanmoins, plusieurs questions demeurent et en premier lieu celle du financement de la donnée. Est-ce que l'usage de ces données vaut l'investissement ? Comment exploiter ces données ? Ainsi, les enjeux globaux relèvent de l'efficacité et de la valorisation de la data mais aussi de sa disponibilité. En effet quelle répliquabilité de méthode, de production et de traitement, alors que les données sont détenues par une entreprise privée ? Il est donc nécessaire de prendre des précautions quant à l'utilisation, à la fois scientifique et politique de ces données.

Les données urbaines ainsi produites sont massives (par exemple, 3,5 milliards de traces numériques fournies par Leboncoin pour l'étude menée sur l'exode urbain, cf. infra) mais hétérogènes dans leurs sources (statistiques, rapports techniques, capteurs), dans leurs formats (types de fichiers différents), dans les médias employés, dans leurs granularités spatiales et temporelles mais aussi dans la légitimité de la source. Le volume de données est tel qu'il est parfois nécessaire de faire appel aux services spécifiques de ceux qui produisent la donnée pour pouvoir la traiter et l'analyser. Dans le cas de l'étude sur l'Exode Urbain, réalisée par Alexandre Coulondre à partir des données du site Leboncoin, la disponibilité de l'équipe de data scientists a été déterminante. L'extraction des données et leur traitement peuvent parfois être faciles parfois plus complexes, en fonction des sources (Google Maps, Uber, Twitter). Comment combiner ces données ? Quel serait alors le langage commun pour les combiner ?

2 Face à l'ubérisation de la donnée, construire sa propre enquête Mobilités

L'utilisation de données produites par d'autres, dans d'autres contextes pose une série de difficultés méthodologiques. Elles peuvent être de nature différente, surtout dans leurs utilisations pour les savoirs spatiaux :

- **Géoréférencement** : dans le cas de l'Etude sur l'exode urbain, les données de localisation des personnes formulant des recherches sont fiables si les recherches sont faites à domicile mais qu'en est-il des recherches effectuées en déplacement ?
- **Représentativité et épaisseur sociologique** : les données fournies ne présentent aucune information socio-démographique. Les populations sont mal définies contrairement à ce qui est fait en statistique conventionnelle. On ne sait pas toujours avec précision qui clique », ce qui explique et légitime les protocoles mixtes.

Les données « classiques » produites avec des méthodes directes, comme celle employée pour l'Enquête Globale Transport (EGT), peuvent se combiner à des données nouvelles (big data) qui sont essentiellement des comptages. Une manière de combiner les données est de les produire en propre. C'est ce que propose l'IPR avec son Enquête Mobilité par GPS (EMG) qui ambitionne de compléter la méthodologie conventionnelle employée dans l'EGT qui existe depuis 1976. L'EMG permet d'expérimenter une enquête autonome qui repose sur une méthodologie en trois étapes : une pré-enquête en ligne pour connaître des éléments de profils et des habitudes de déplacement, une collecte de données GPS sur 7 jours consécutifs effectuée avec des traceurs GPS spécifiques (un point toutes les deux secondes) et un journal de bord à remplir par chaque participant. A l'analyse automatisée des points GPS sont associées des post-enquêtes téléphoniques qui permettent de compléter les motifs de déplacement et vérifier les modes détectés par les algorithmes.

La production et la vente de données massives par des acteurs privés entraînent ce qui peut être qualifié d'ubérisation de la donnée qui n'était pas faite pour être traitée et analysée. La sphère publique a quelque peu perdu la main sur cette production qui paraît à la fois lourde et coûteuse. Toutefois, une mutualisation des coûts pourrait être envisagée pour assurer une production de données publiques, nettoyées et produites en vue d'être analysées, d'autant plus que les modèles de données sont souvent internationaux. Comment faire entrer ces modèles en pertinence avec les territoires ? Le modèle de la qualité de l'air ou celui du calcul de trafic sont élaborés sur des modèles nord-américains. Comment les appliquer aux villes européennes ? Et par extension, comment prévoir la réception politique de ce type de données.



MODALITÉS

Enquête sur 20 semaines en 2023, 167 participants par semaine. 620 traceurs réutilisables.

BD de base de GPS. 500 millions de point renseignés par profil de la personne.

L'enquête téléphonique finale de réajustement a un coût pour valider des informations (indispensable car les algorithmes ne sont pas totalement fiables)

L'incitation financière sous forme de bon d'achat a permis de fidéliser des enquêtés.e.s

Restructurer l'enquête publique : le cas de l'enquête régionale sur la mobilité des Franciliens

BILAN

Une enquête immédiatement exploitable avec un nettoyage limité de la donnée

Recrutement à améliorer : ouvriers et chômeurs difficile à enquêter

Réduction du coût unitaire de 2,5 par rapport à l'EMC²

3 Big Data et action publique

MyTraffic, un retailer, offre un outil pour l'Observatoire de l'Offre Commerciale de la Métropole du Grand Paris.

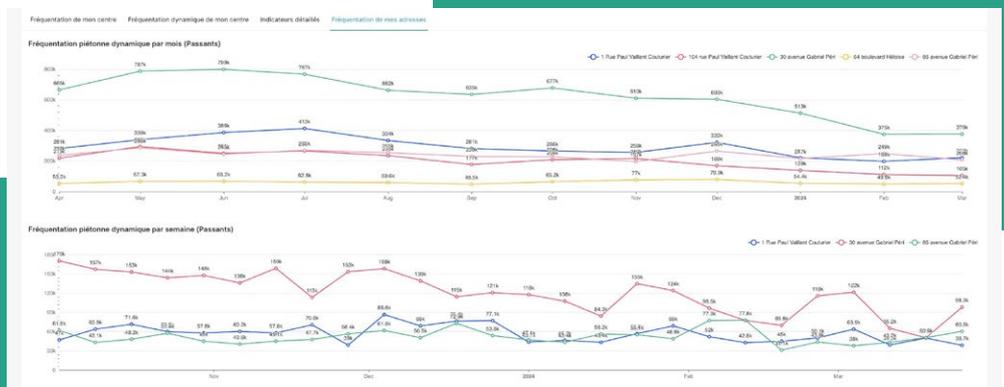
Quels seraient les effets de ces données sur l'aménagement et les politiques publiques ? Peuvent-elles induire un changement de regard pour faire changer les manières d'agir ? Une question essentielle se pose : comment passer de la donnée brute à la prise de décision ?

Sur la question des linéaires commerciaux et des grands centres commerciaux par exemple, l'enjeu est de créer un centre de ressource qui permet d'accompagner les installations commerciales, aussi bien en termes d'ingénierie de projet que d'accompagnement juridique. La MGP propose d'accompagner les communes pour une co-construction de projets avec des offres de service et d'animation, tout en travaillant la maîtrise foncière, avec l'aide de la Foncière Centre-Ville Vivant par exemple. Pour mener à bien cet accompagnement, des données, de fréquentation notamment, sont nécessaires. C'est pour cette raison qu'en 2020 est mis en place l'observatoire de l'attractivité des centres-villes. Il s'agit d'un observatoire de l'offre commerciale co-construit avec la Chambre de Commerces et d'Industrie Paris Ile-de-France.

Comment évoluent les flux dans la métropole sur différents périmètres de centre-ville ? Comment caractériser et expliquer les effets d'évolutions contextuelles et d'évènements ponctuels (chantiers, travaux, rencontres) sur l'activité commerciale. ? Comment éviter la vacance commerciale et la mono-activité commerciale ? MyTraffic, un retailer, offre un outil pour cet observatoire et les missions qui sont les siennes. Si son activité était initialement privée et se focalisait sur l'évaluation de centre commercial par exemple, elle propose ses services aux collectivités depuis 2 ans. Malgré tout, ce type d'entreprise, contrairement aux instituts, n'a pas de culture territoriale ni de compréhension des métiers pour l'analyse de ce type de données.

Des cultures professionnelles différentes se confrontent alors avec l'ouverture d'une question nouvelle : celle de l'altruisme de la donnée. Cela suppose-t-il un échange de bons procédés ? Cette modalité, qui diffère du simple rapport marchand, soulève d'autres interrogations éthiques.

Centres Villes Vivants, Extrait du rapport d'analyse trimestriel des flux de piétons dans la commune d'Argenteuil, d'après des données MyTraffic, 2024



4

Reconfiguration métropolitaine et urgence de la donnée

La temporalité des enquêtes est centrale lorsqu'il s'agit de répondre à l'actualité, notamment concernant les mobilités.

Si les données numériques massives présentent un biais cognitif majeur lié à la fraîcheur de la donnée et à l'IA (effets « wahou » notamment auprès des élus), il convient de rappeler des arguments techniques et méthodologiques souvent peu audibles. La question stratégique de la donnée propre se pose avec plus de force encore. Elle témoigne d'une évolution importante. Avant les données statistiques étaient basées sur l'offre alors que maintenant les données sont basées sur la demande, à laquelle il conviendrait de répondre instantanément.

Cette question de la temporalité des enquêtes est centrale lorsqu'il s'agit de répondre à l'actualité, notamment concernant les mobilités. L'enquête ménage déplacement, par exemple, a lieu tous les 8-9 ans. La dernière en Ile-de-France, dont les résultats sont disponibles, date de 2018. La toute dernière est en cours mais les résultats ne seront disponibles qu'en 2028. En 2021, la base de données de l'EGT 2018 n'avait pas été transmise alors que des questions se posaient avec insistance :

- Quelles évolutions des mobilités en contexte post-covid ?
- Etude de la relation origine – destination sur le boulevard périphérique ?
- Quelle part modale du vélo par département ?
- Comment distinguer les mobilités en fonction des jours de semaine ?

Pour aller plus loin...

Enquête régionale sur la mobilité des Franciliens

<https://www.institutparisregion.fr/mobilite-et-transport/deplacements/enquete-regionale-sur-la-mobilite-des-franciliens/> et <https://www.institutparisregion.fr/mobilite-et-transport/deplacements/tableau-de-bord-de-la-mobilite-en-ile-de-france/>

L'écosystème des transports en commun en Ile-de-France :

<https://www.institutparisregion.fr/nos-travaux/publications/lecosysteme-des-transport-en-commun-en-ile-de-france/>

Territoires cyclables, étude de l'Institut Paris Région

<https://www.institutparisregion.fr/mobilite-et-transport/modes-actifs/territoires-cyclables/>

Les territoires de la marche en Ile-de-France :

<https://www.institutparisregion.fr/nos-travaux/publications/les-territoires-de-la-marche-en-ile-de-france/>

Le mass transit à l'heure du télétravail et de la sobriété énergétique :

https://www.institutparisregion.fr/fileadmin/NewEtudes/000pack3/Etude_2844/NR_958_web.pdf

Comment alors répondre à ces questions sans avoir les données actuelles ? Tout l'enjeu de l'étude sur l'Exode Urbain résidait dans le changement de ces temporalités de recherche et de production de la donnée. Comment mesurer un effet Covid dès 2021 ?

Le réflexe initial est de travailler sur des données publiques. En l'occurrence, ce type de données ne pouvaient pas être mobilisées parce qu'il présente 4-5 ans de décalage. Les données fournies par Leboncoin avaient notamment plusieurs mérites : elles étaient disponibles et fraîches, elles répondaient pleinement à la question et elles étaient gratuites (mais pour combien de temps ?). Si elles permettent de dresser rapidement un portrait d'une situation à un instant précis, les données dérivent. La structure d'un site change au cours du temps. Les données ne sont alors pas comparables d'une période à une autre.

CONCLUSION

Les données numériques (big data, données retail), combinées à des enquêtes ciblées et menées par des organismes d'État, renouvellent les possibilités de captation de la donnée, en particulier à l'échelle fine des individus. Elle permet, au-delà des déplacements, d'interroger des comportements de mobilités mais aussi la manière dont les politiques publiques peuvent, à partir de ce nouveau matériau analysé, agir et proposer un gouvernement des conduites individuelles dans le cadre des transitions écologiques. Entre foisonnement de la donnée, le coût induit par son traitement et protection des individus au nom du RGPD, c'est un renouvellement des registres de l'action qui s'ouvre pour le territoire grand-parisien dans la manière d'appréhender son fonctionnement en archipel (accompagnement des flux et des logiques d'ancrage).

L'équipe

Directeur scientifique : [Daniel Béhar](#)

Responsable de l'axe 1 : [Claire Aragau](#)

Responsable de l'axe 2 : [Hélène Dang Vu](#)

Responsable de l'axe 3 : [Julien Aldhuy](#)

Animatrice scientifique : [Léa Donguy](#)