



**HAL**  
open science

## La mise en données des villes : entre fascination technique et innovation urbaine

Boris Mericskay

► **To cite this version:**

Boris Mericskay. La mise en données des villes : entre fascination technique et innovation urbaine. Atelier “ Mesure et démesure ”, UMR PASSAGE, Nov 2023, Bordeaux, France. halshs-04473275

**HAL Id: halshs-04473275**

**<https://shs.hal.science/halshs-04473275v1>**

Submitted on 22 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# La mise en données des villes: entre fascinations techniques et innovations urbaines

# Des villes et des données

- Les villes ont toujours été le terrain privilégié d'expérimentations et de développements de systèmes informatiques utilisant des données
  - Des premières BDU aux jumeaux numériques en passant par les SIG urbains, les usages comme le statut de l'information territoriale dans l'action publique n'ont cessé d'évoluer et de se développer depuis 50 ans
- La mise en données / géonumérisation des villes
  - Communications, achats, déplacements, avis, procédures administratives...la plupart des activités urbaines donnent désormais lieu à une production massive de données numérisées à composante spatiale



# Transition numérique de l'action publique

- Au cœur de la relation entre ville et numérique, les données transforment les manières d'habiter, de gérer et d'aménager
  - Produites de manière exponentielle par de nombreux acteurs elles inondent tous les champs de l'urbain et viennent renouveler les pratiques des pouvoirs publics, des entreprises comme des chercheurs
  - D'une ressource informationnelle au service de la connaissance, les données sont devenues une infrastructure essentielle de l'action publique largement mise en avant comme leviers d'amélioration, d'innovation ou d'évaluation des politiques et des services urbains



REVUES NUMÉRIQUES NEWSLETTERS RECHERCHE ABONNEZ-VOUS

**la gazette**  
Le magazine des collectivités et des régions

ACCUEIL > SMART CITY : LES CLÉS DE LA VILLE INTELLIGENTE > Smart city : rassembler les données en vue de piloter la ville

DONNÉES INTELLIGENTES

## Smart city : rassembler les données en vue de piloter la ville

Publié le 06/02/2018 • Par **Baptiste Cessieux** • dans : **Dossiers d'actualité** • France



La ville intelligente est bien plus qu'un énième objet « smart », mais elle répond aux mêmes règles : récupérer un maximum de données afin d'automatiser les tâches et d'affiner la prise de décision.

Marylin Gobert / La Gazette

RÉAGIR À CET ARTICLE

SUR LE MÊME SUJET

Les villes moyennes aussi, ont droit à la smart city

Les Français plébiscitent l'administration en ligne et la smart city

Smart city : « L'hyper-technicisation laisse peu de place au citoyen »

La smart city, à quel prix ?

INSTITUT Montaigne

Cohésion sociale ▼ Compétitivité économique ▼ Efficacité de l'État ▼ Coopérations internationales ▼

Rechercher un rapport, une publication, un expert...

Accueil > Villes et territoires > **Villes, à vos données !**

RAPPORT Juillet 2021

## Villes, à vos données !

Villes et territoires Technologies Action publique

↓ TÉLÉCHARGER PARTAGER



L'USINENOUVELLE

Secteurs ▼ Obsessions ▼ Rendez-vous ▼ Régions ▼ Vidéos & Podcasts ▼

RECHERCHE MAGAZINE NEWSLETTERS

## Smart City : quand les données irriguent la ville

En s'imposant au cœur des villes intelligentes, les données ouvrent la voie à de nouveaux services améliorant l'écosystème urbain. Enjeu de pouvoir, elles représentent un marché d'avenir qui attire les convoitises. Collectivités, délégataires de service public et géants du numérique sont sur les rangs, tandis que l'open data fait valoir sa transparence.

Réservé aux abonnés

Alexandre Coutoacouto@industrie-technologies.com

IT

28 décembre 2019 10h00

6 min. de lecture



MINISTÈRE DE LA TRANSFORMATION ET DE LA FONCTION PUBLIQUES

Direction interministérielle de la transformation publique

Espace presse Boîte à outils

Rechercher

Transformer l'action publique Améliorer l'expérience usagers Diffuser l'innovation publique Associer les citoyens L'offre d'accompagnement de la DITP

Accueil > Boîte à outils > Outils et formations > Les données publiques, un outil pour les services publics ?

Webinaire

## Les données publiques, un outil pour les services publics ?

Publié le 10 septembre 2021 par Mission « Innovation publique »

Innovation publique Transformation numérique

Retour sur l'épisode 6 des "Mardis de l'Innovation", organisé par la DITP le 7 septembre dernier, avec les regards croisés d'experts de la valorisation des données : Audrey Moisan, Chef de projet, Mission Innovation de la Direction interministérielle de la transformation publique ; Mathilde Hoang, Head of open data and data sharing chez Etalab (Direction interministérielle du numérique), Cécile Le Guen, Cheffe de projets Data Science à la DITP et Louis Rached Responsable Innovation et Digital, Ville d'Issy-les-Moulineaux.



ADMINISTRATEUR GÉNÉRAL DES DONNÉES

## LES DONNÉES AU SERVICE DE LA TRANSFORMATION DE L'ACTION PUBLIQUE

RAPPORT AU PREMIER MINISTRE SUR LA GOUVERNANCE DE LA DONNÉE 2015

MISSION BATHOREL

## Pour une politique publique de la donnée

DECEMBRE 2021

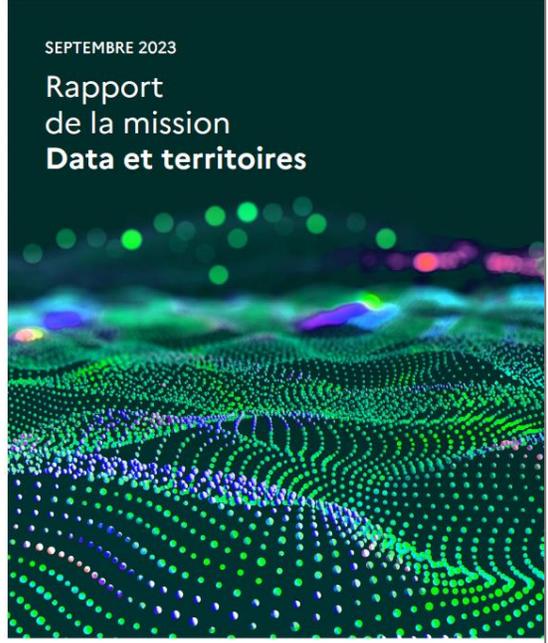


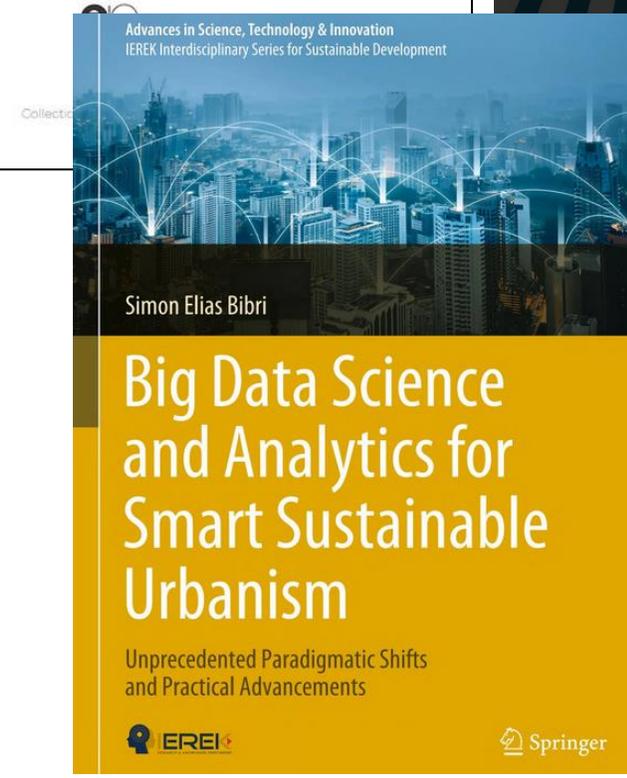
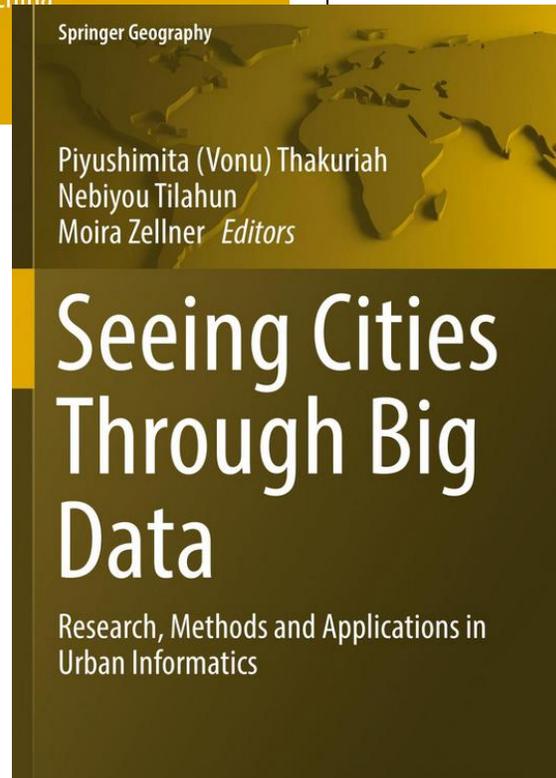
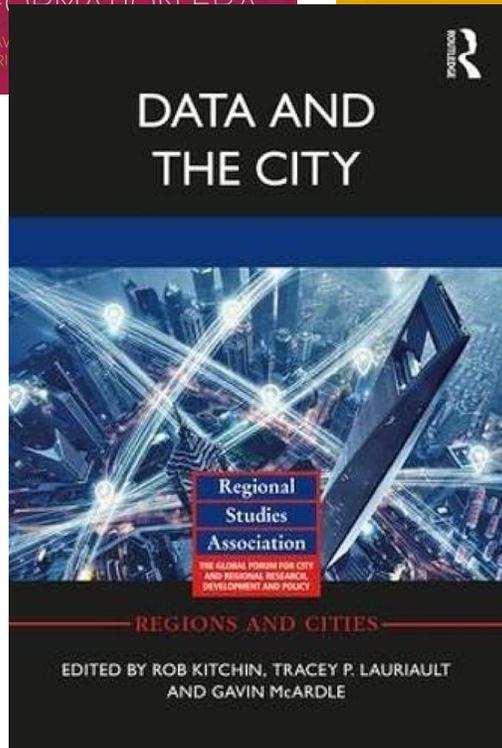
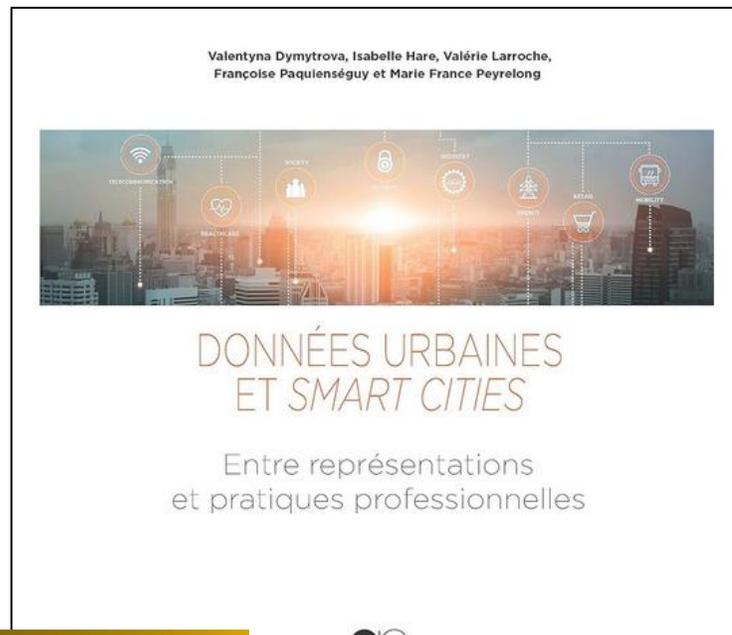
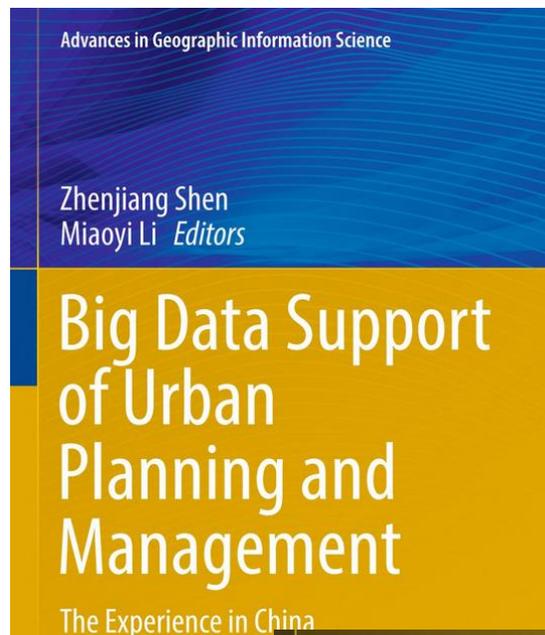
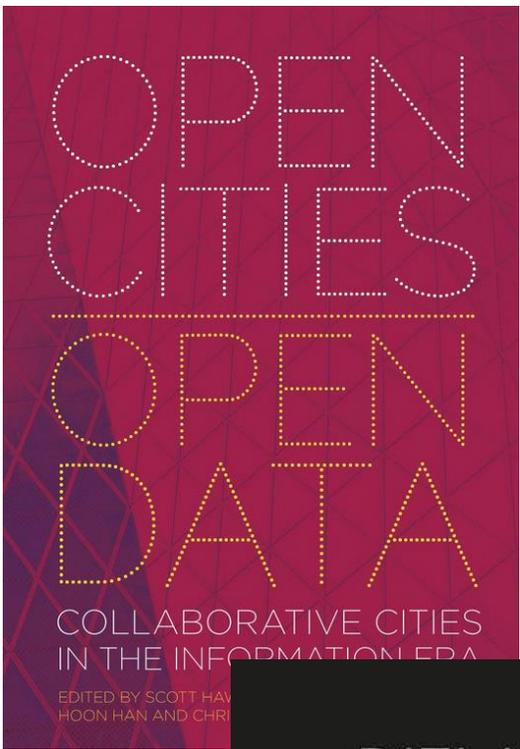
MINISTÈRE DE LA TRANSFORMATION ET DE LA FONCTION PUBLIQUES

Liberté Égalité Fraternité

SEPTEMBRE 2023

## Rapport de la mission Data et territoires



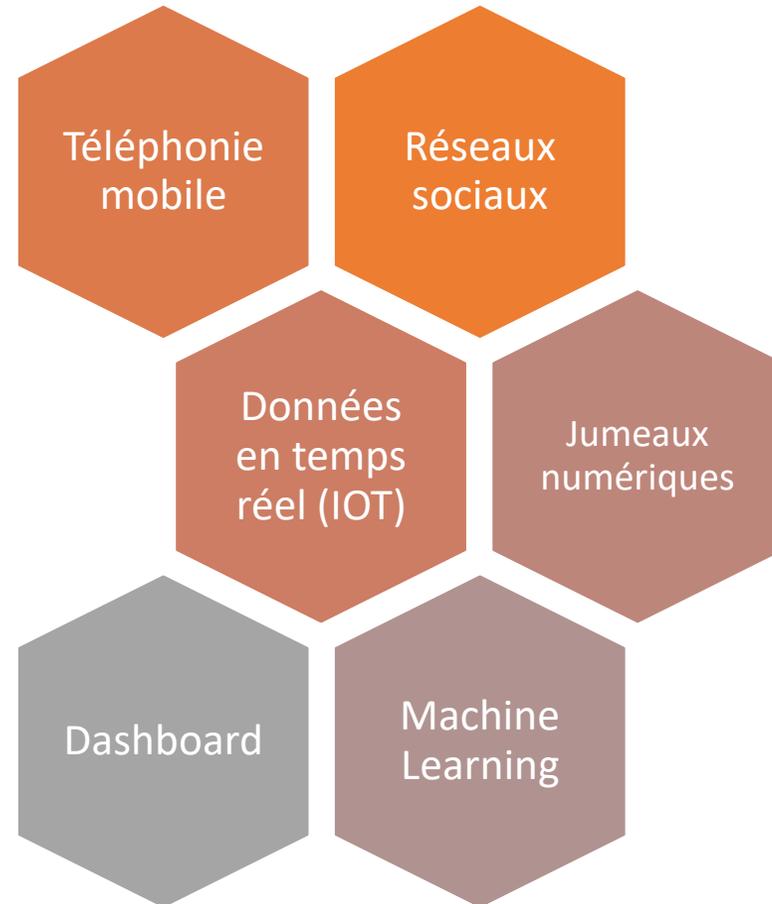




# Fascination <> Innovation

- **Fascination** où la dimension techniciste de la mise en données des villes provoque un effet de « sidération »
  - **Injonction** et **incantation** du pouvoir des données à mieux comprendre, piloter et optimiser les villes pour les praticiens
  - Enthousiasme sur les potentialités analytiques des données du côté des chercheurs pour saisir de nouvelles logiques à des échelles fines
- Cette **fascination** autour de la mise en données des villes se traduit-elle par la **créativité** et **l'innovation** dans les pratiques opérationnelles comme de recherche ?

# Explorons quelques serpents de mer

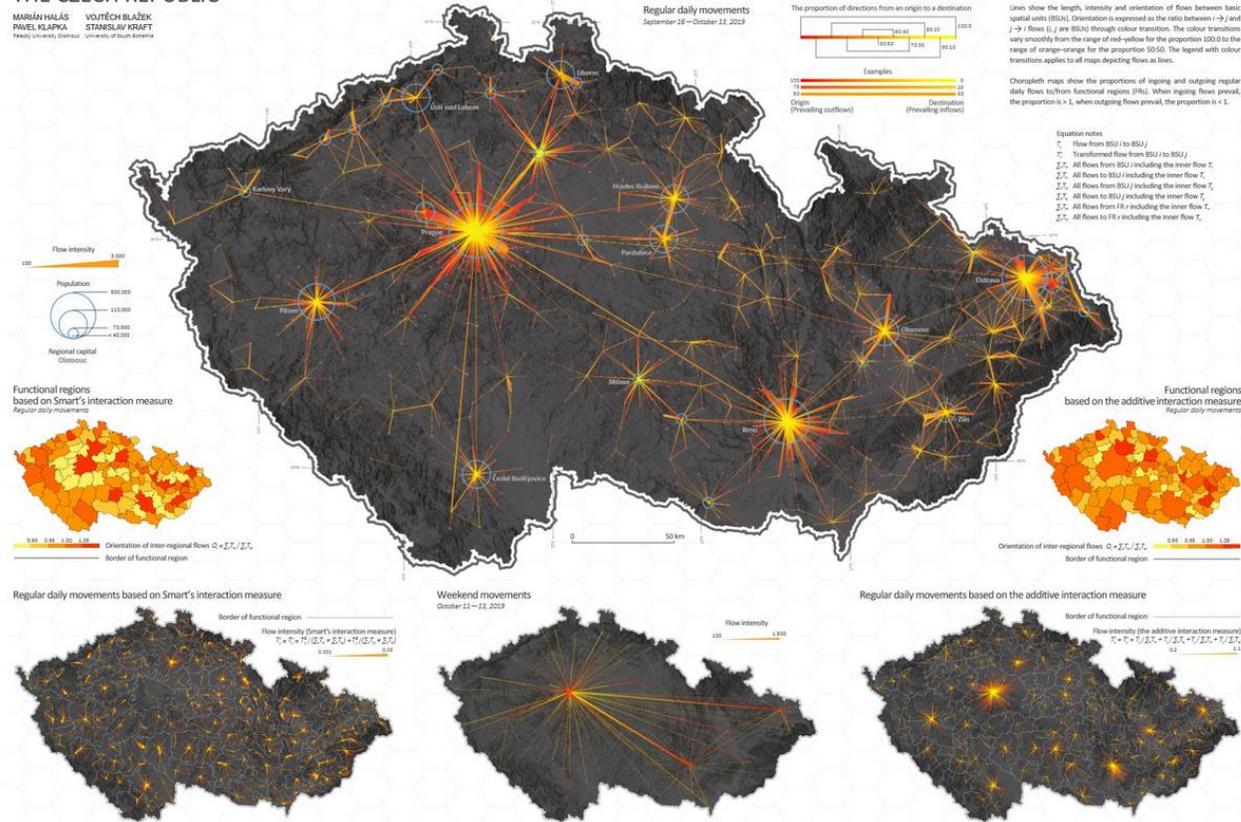




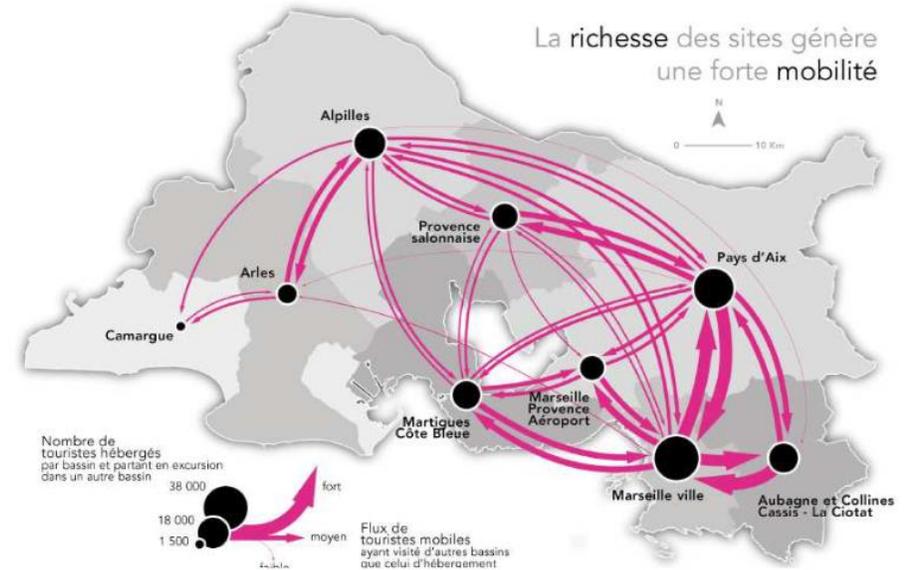
# Réalité technique

## POPULATION MOVEMENTS BASED ON MOBILE PHONE LOCATION DATA THE CZECH REPUBLIC

MARIÁN HALÁŠ VŮTĚCH BLÁŽEK  
PÁVEL KLAPKA STANISLAV KRAFT  
Masaryk University of Brno Faculty of Science



## LA MOBILITE TOURISTIQUE



Business Services

Zone d'activité / nuitée des visiteurs

Distribution des visiteurs par zones de provenance

DREUX/GARE	10.2%
DREUX/ZONÉ ACTIV.	9.3%
DREUX/FALAISE	4.3%
DREUX/HARICOT	3.3%
DREUX/BATES/UEST	3.2%
HOUDAN	2.8%
VERNOUILLET/CENTR.	2.3%
CHERISY	2.3%
GARNAY	2.1%
MESNIL-SUR-L'ESTREE	2.1%
MEZERES-EN-DROUAIS	1.7%
ABONDANT	1.5%
SAINT-LUBIN-DES-JON.	1.4%
TREMBLAUX-LES-VILLAG.	1.0%
TREON	0.9%
SAINT-REMY-SUR-AVRE	0.9%

Cartographie des zones de provenance

Rechercher un lieu dans la carte

Date: Y Q M W D, avr. 1 2019 - déc. 19 2019

Jour de la semaine: lundi, mardi, mercredi, jeudi, vendredi, samedi, dimanche

Catégorie Visiteurs: Etrangers, Français

Durée:  < 30 min,  30 min à 3h,  > 3h

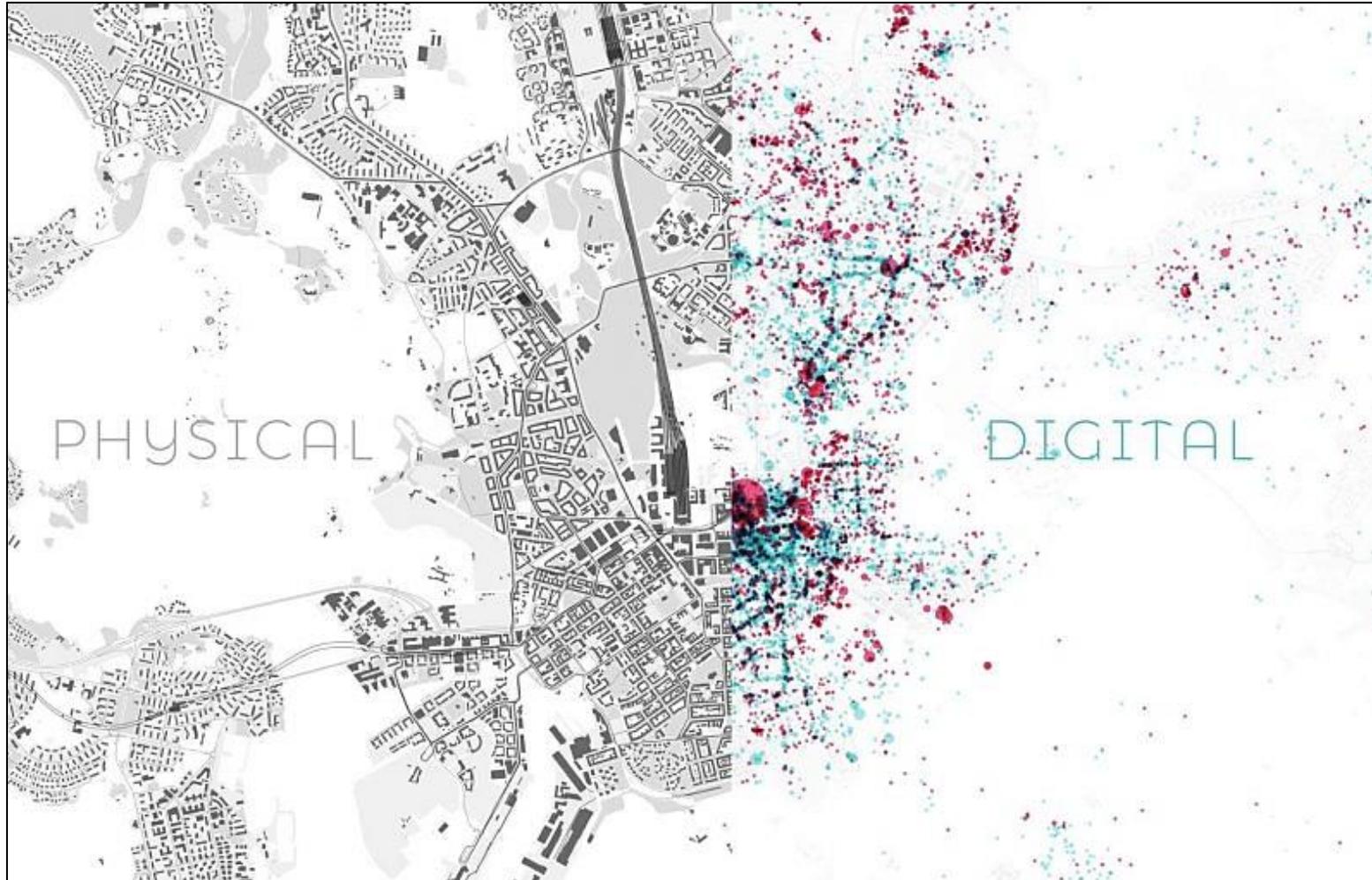
Zone de chalandise:  Zone d'activité,  Zone de nuitée

Zonage:  Bassin de vie,  Communes,  Iris

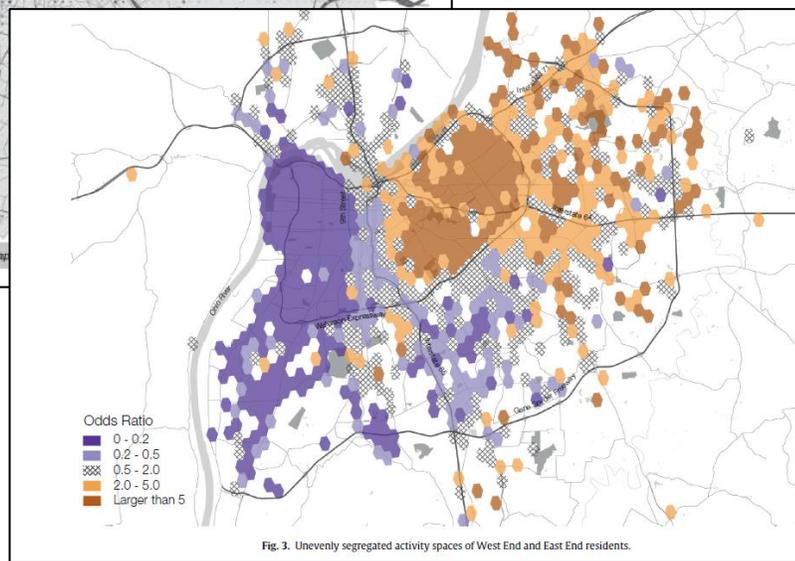
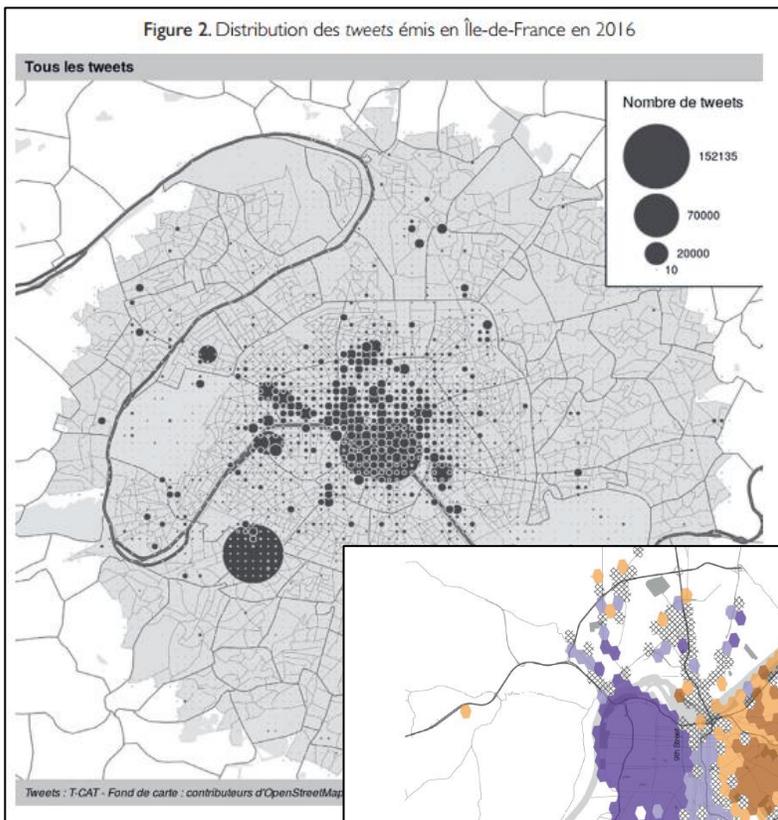
Zone d'analyse: Centre-ville restrei...

Le lieu de nuitée la veille (résidence ou hébergement) référence le lieu de présence majoritaire entre 00h et 06h quand le lieu d'activité principal en journée (lieu de travail, ...) est identifié par le lieu de présence majoritaire entre 09h et 17h.

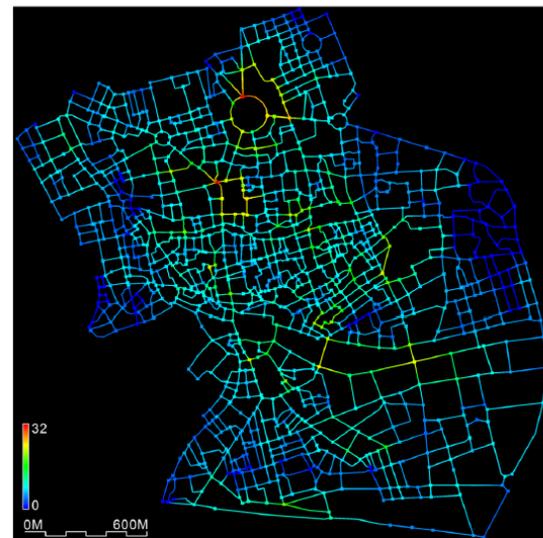
# Fascination technique



# Réalité technique

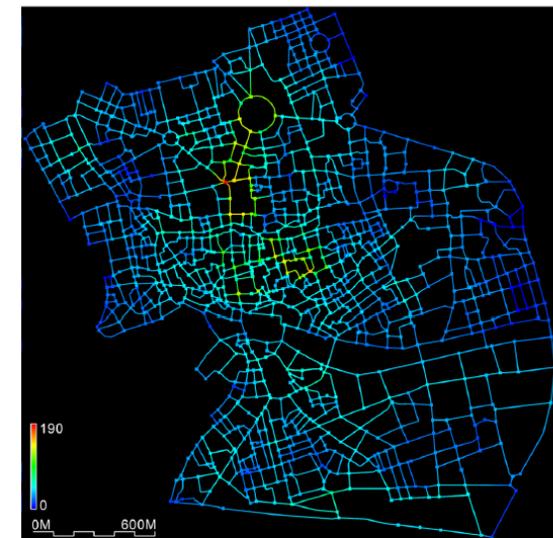


Foursquare



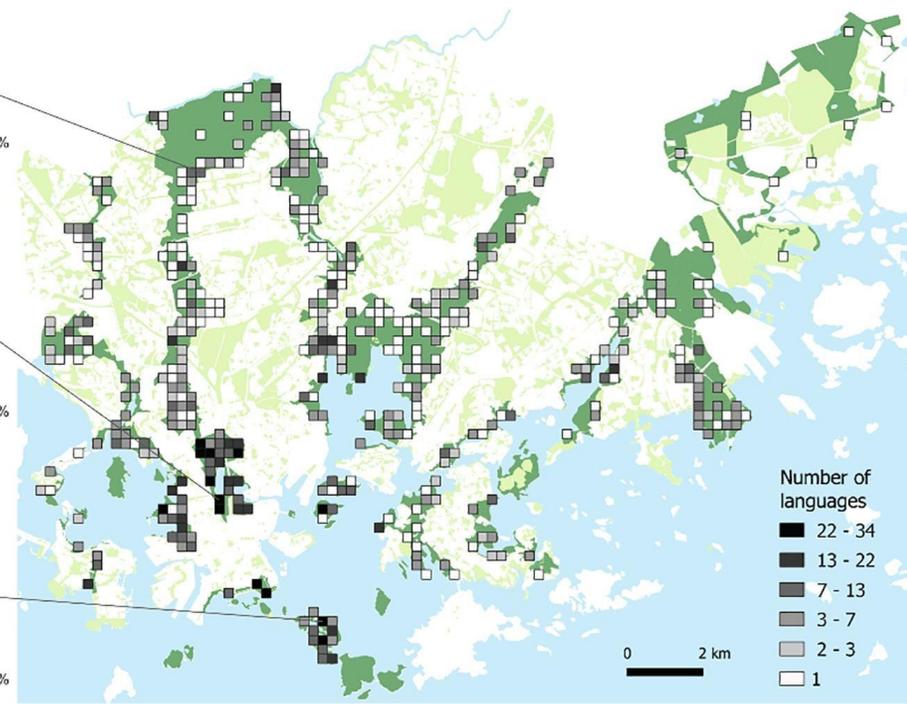
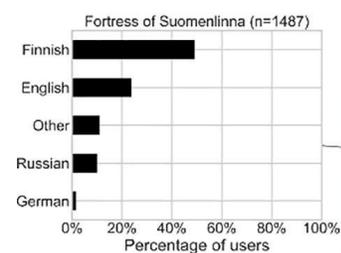
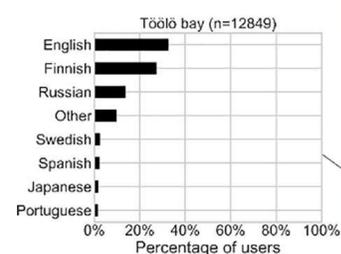
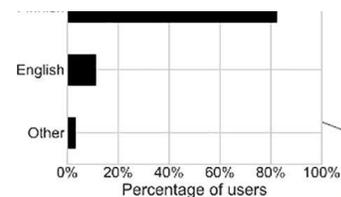
(a)

Google Places



(b)

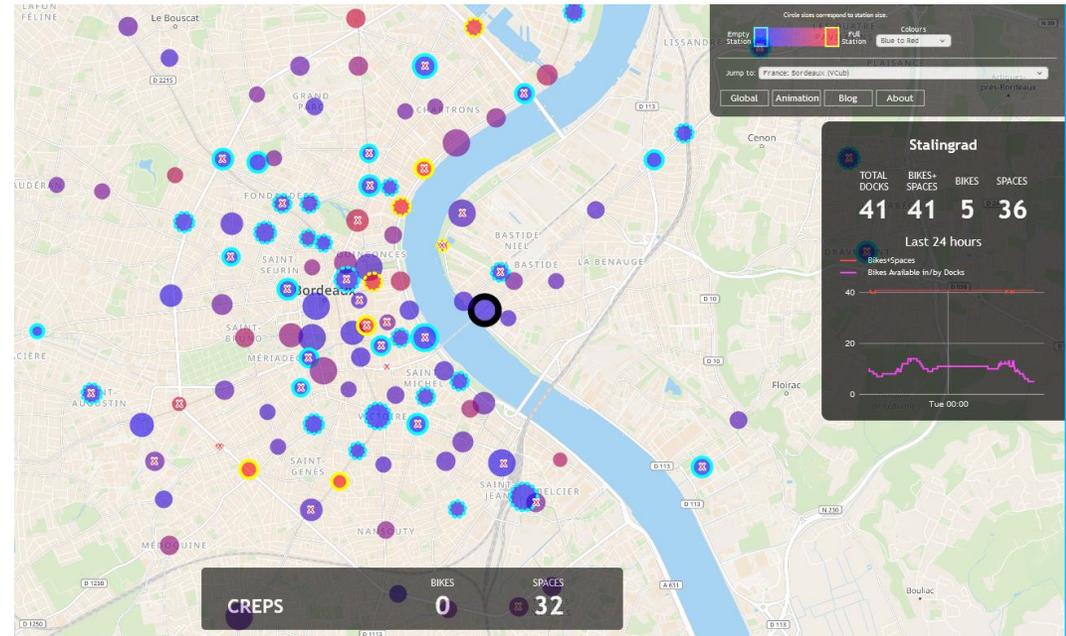
Visualization by network.



# Fascination technique



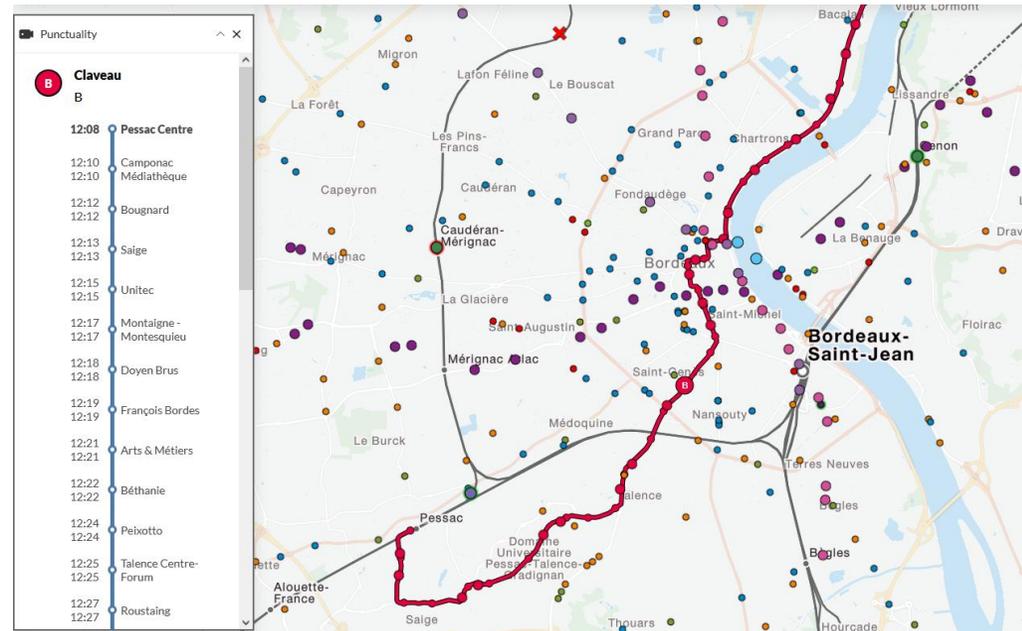
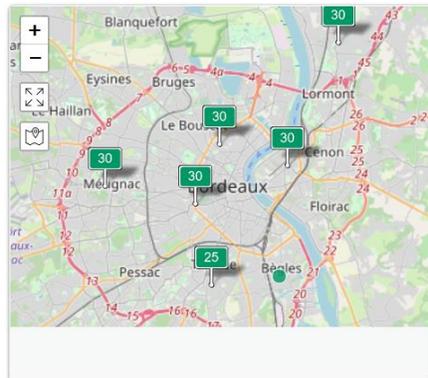
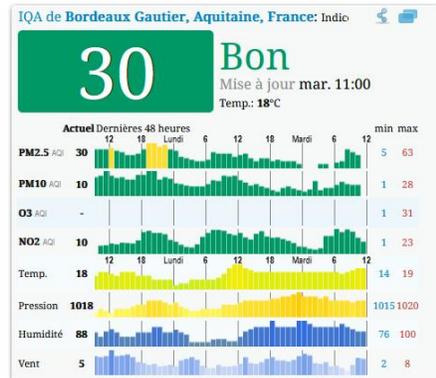
# Réalité technique



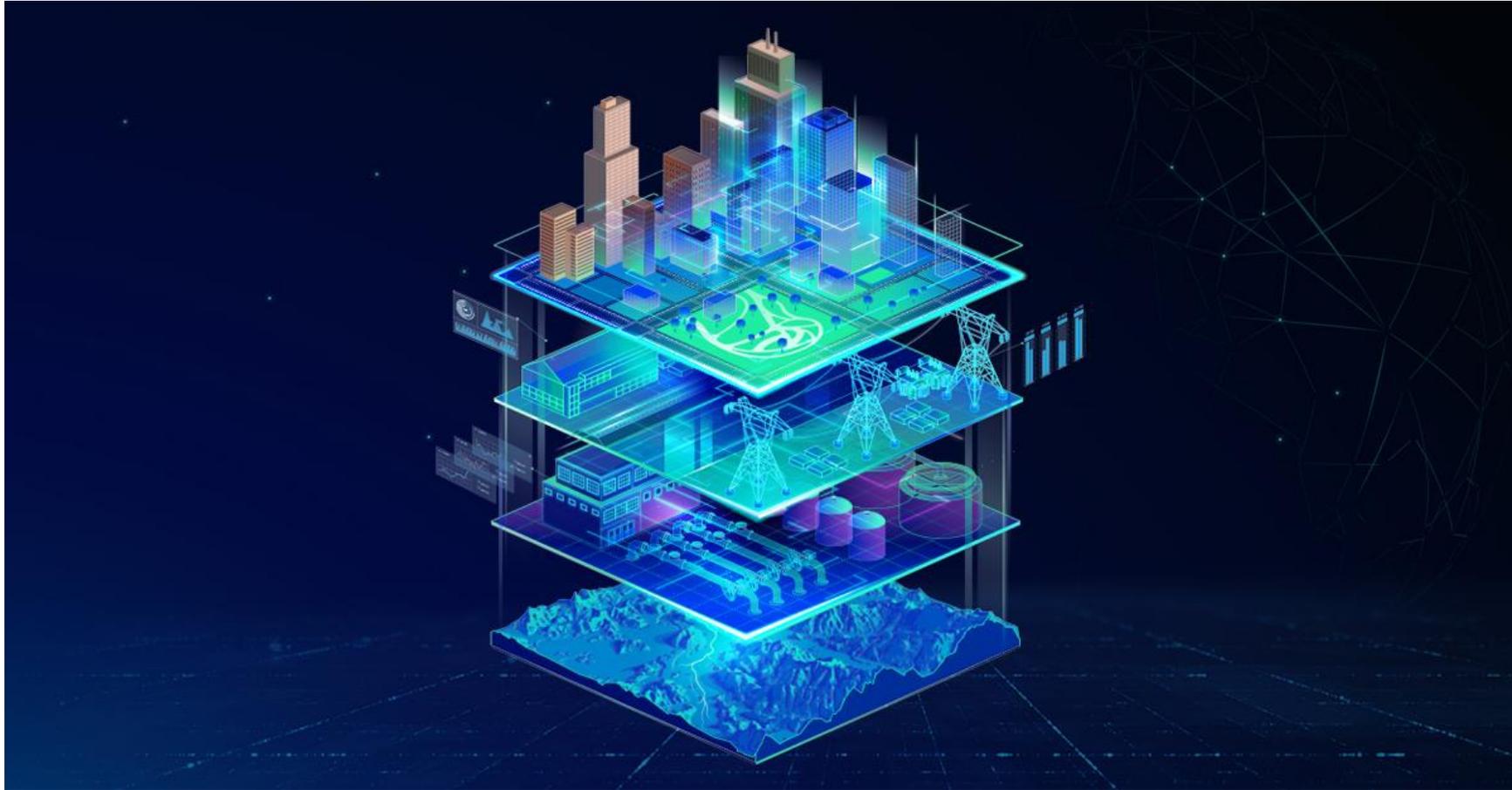
## Pollution de l'air à Bordeaux

### Indice de la qualité de l'air (IQA) en temps réel

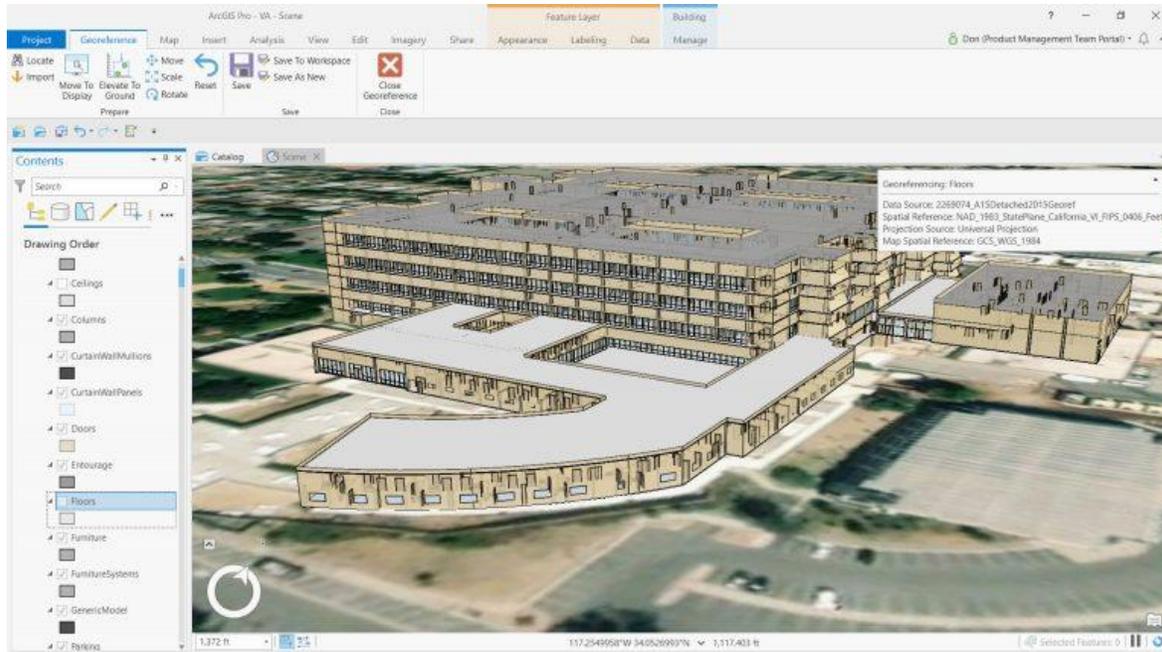
BORDEAUX GAUTIER, GRAND PARC, BASTIDE, MERIGNAC, BASSENS, TALENCE, LOCALISER LA VILLE LA PLUS PROCHE, RECHERCHER VOTRE VILLE  
 Lyon Centre  
 13 - Bon - mar., 11AM  
 Il semble que vous ne soyez pas situé dans Bordeaux Gautier, Aquitaine, et que la ville la plus proche de votre position soit Lyon. Voulez-vous voir l'indice de la qualité de l'air pour Lyon Centre ?  
 ok Non merci.



# Fascination technique



# Réalité technique

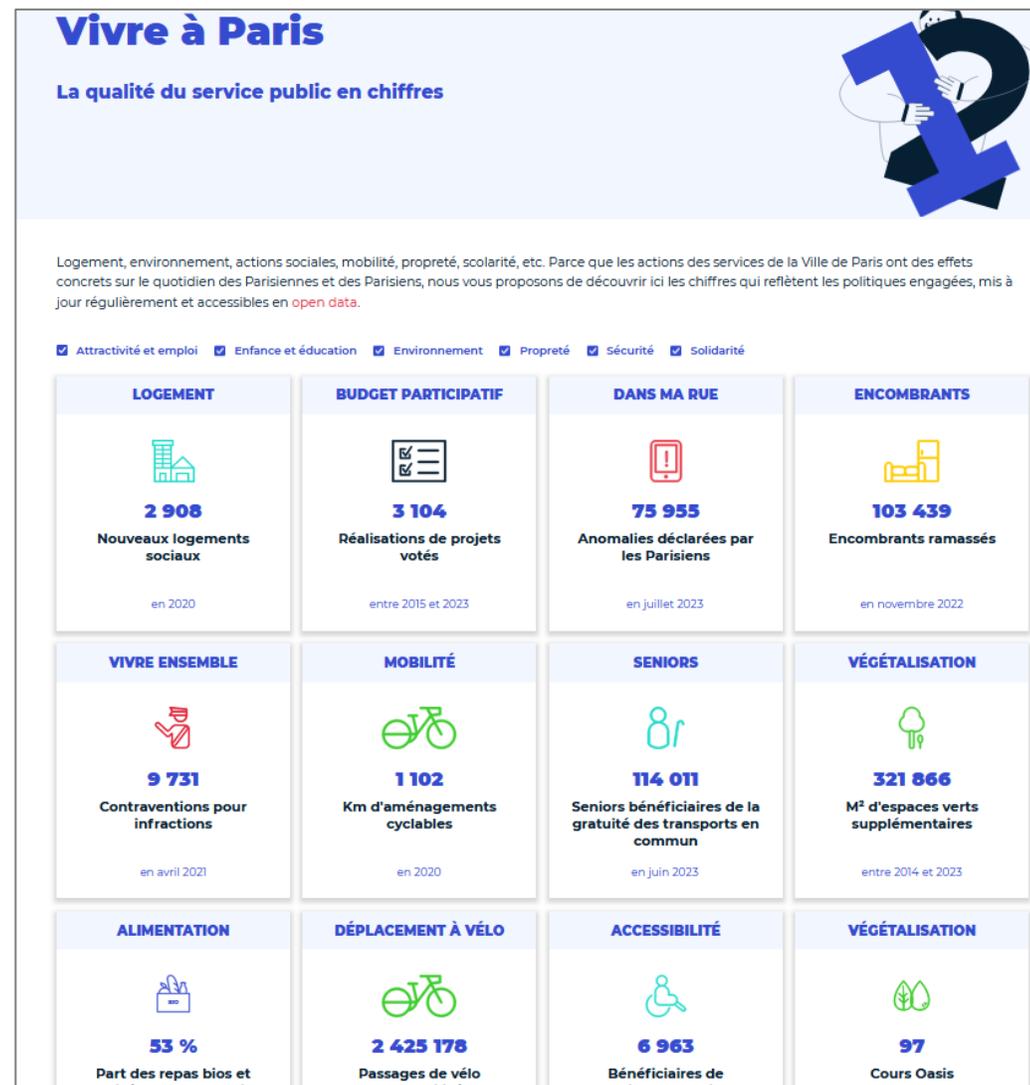
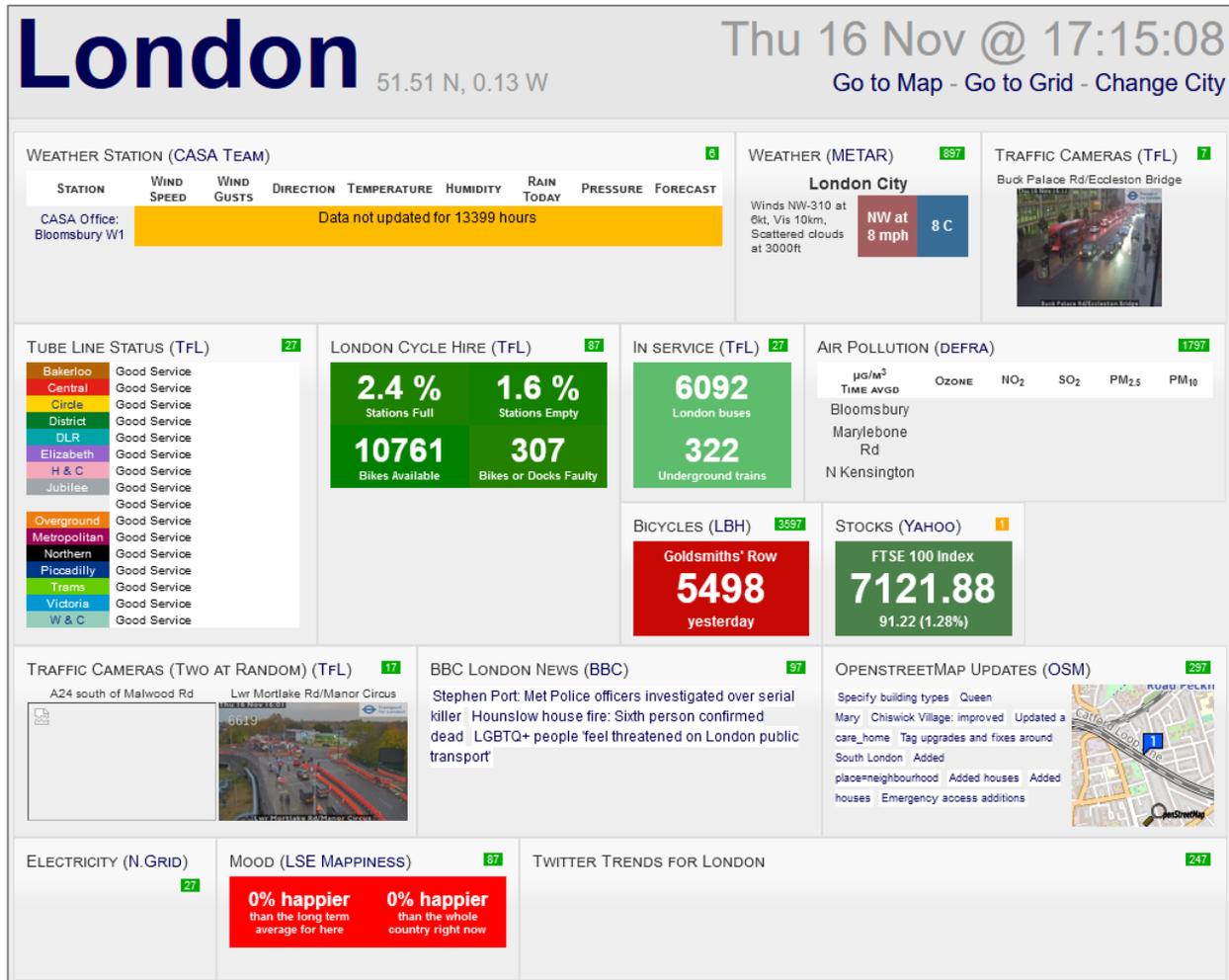


# Fascination technique

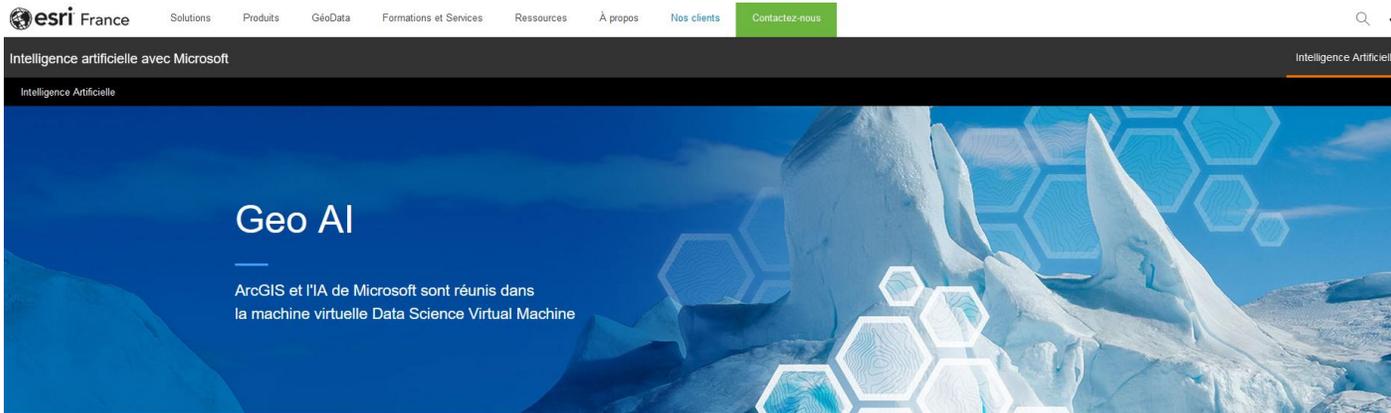


Pilotage des villes par les données pour **quantifier, suivre et évaluer**

# Réalité technique

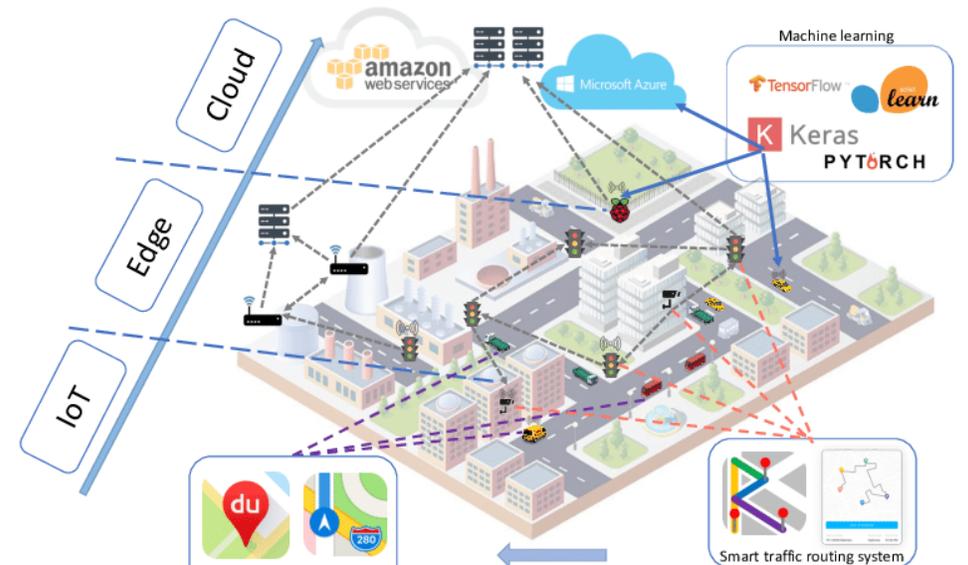


# Fascination technique

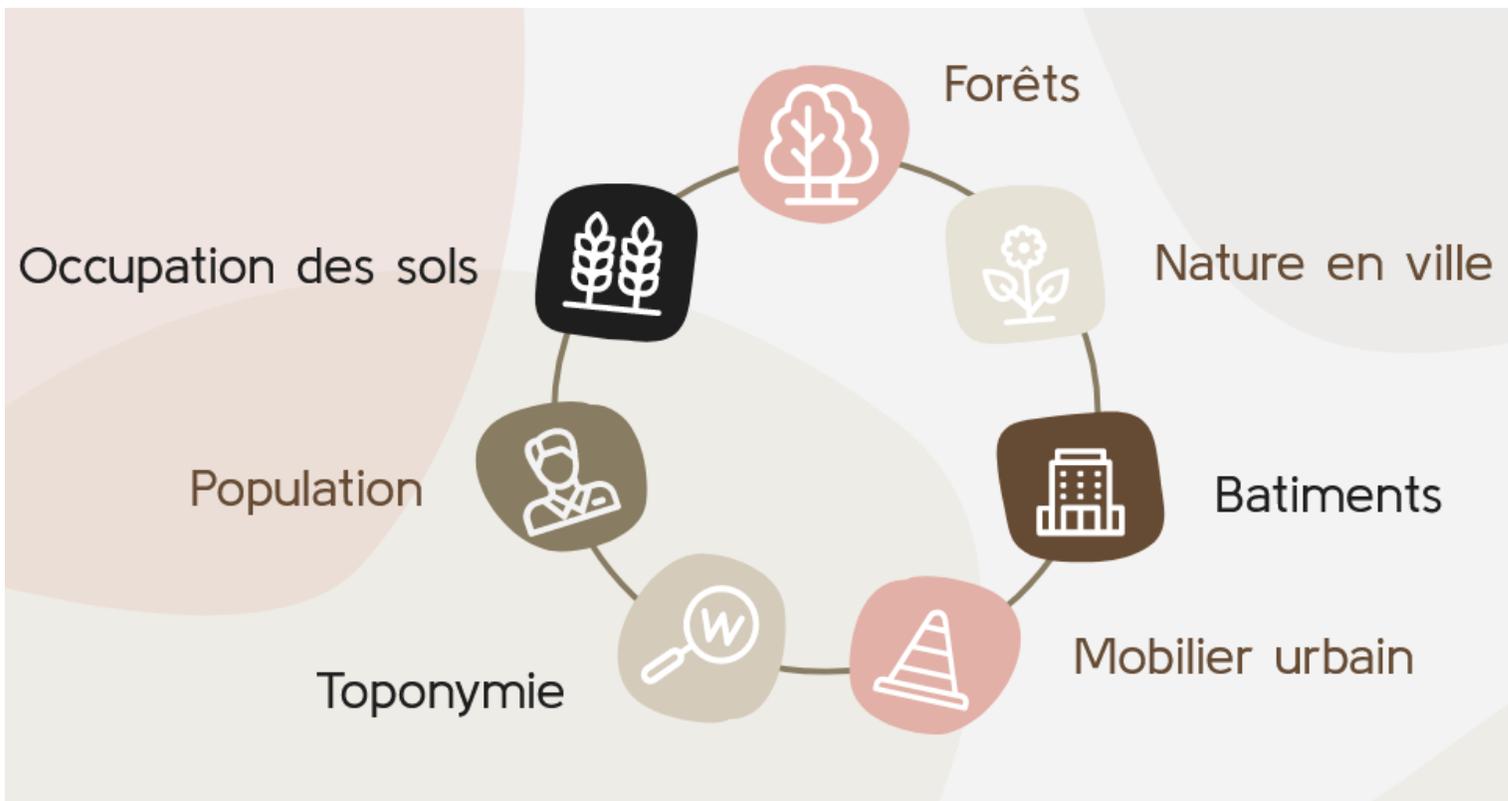


Des perspectives plus approfondies et une prise de décision plus rapide grâce au SIG et à l'IA

Esri et Microsoft ont uni leurs forces pour apporter la puissance de la technologie géospatiale et de l'intelligence artificielle (IA) dans la machine virtuelle Microsoft Data Science (DSVM) sur la plateforme cloud Azure. Geo AI associe les capacités analytiques d'ArcGIS Pro, le SIG bureautique de nouvelle génération d'Esri, à la technologie AI de Microsoft et à ArcGIS Enterprise sur Azure pour fournir des capacités de Geo AI basées sur le cloud.



# Réalité technique



**Apprentissage supervisé : Détection, extraction, segmentation et classification d'objets**

## ChatGPT

Pour créer une carte qui représente le nombre de passages par arrêt, vous pouvez utiliser la bibliothèque `geopandas` pour manipuler des données géospatiales et `folium` pour générer des cartes interactives. Assurez-vous d'installer ces bibliothèques en utilisant la commande `pip install geopandas folium`.

Voici un exemple de script Python qui utilise ces bibliothèques pour créer une carte :

```
python
Copy code

import pandas as pd
import geopandas as gpd
import folium
from zipfile import ZipFile
from io import BytesIO
import requests

# Télécharger le fichier GTFS
url = "https://data.toulouse-metropole.fr/api/v2/catalog/datasets/tisseo-gtfs/files/1"
response = requests.get(url)
zip_file = ZipFile(BytesIO(response.content))

# Charger les données des arrêts
stops_df = pd.read_csv(zip_file.open('stops.txt'))

# Charger les données de la géométrie des arrêts (shapes.txt)
shapes_df = pd.read_csv(zip_file.open('shapes.txt'))

# Fusionner les données pour obtenir les coordonnées des arrêts
stops_with_geometry = pd.merge(stops_df, shapes_df, on='stop_id')

# Créer une GeoDataFrame à partir du DataFrame
gdf = gpd.GeoDataFrame(stops_with_geometry,
                       geometry=gpd.points_from_xy(stops_with_geometry['shape_pt_lon',
                                                    'shape_pt_lat']))

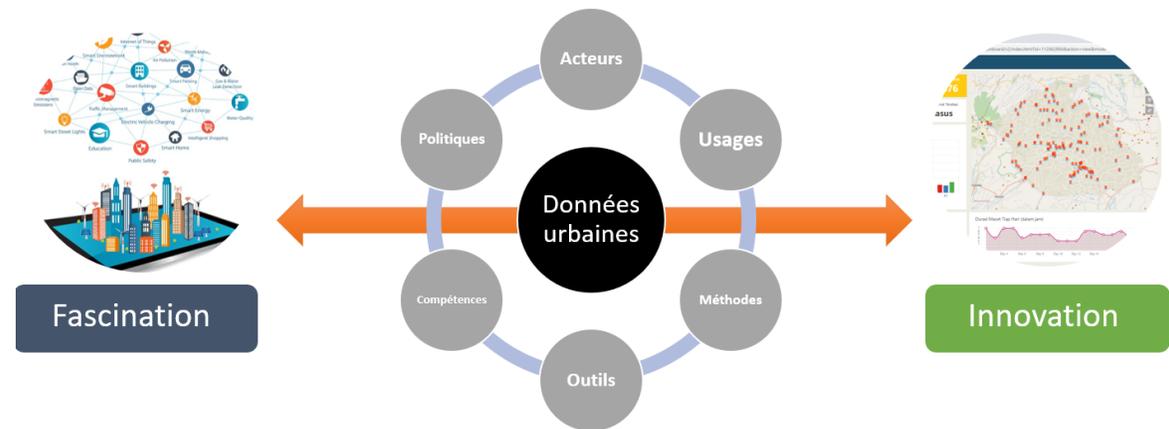
# Compter le nombre de passages par arrêt
passages_par_arrêt = gdf.groupby('stop_name').size().reset_index(name='passages')

# Créer une carte avec folium
m = folium.Map(location=[43.6047, 1.4442], zoom_start=12) # Coordonnées pour Toulouse

# Ajouter des marqueurs pour chaque arrêt
```

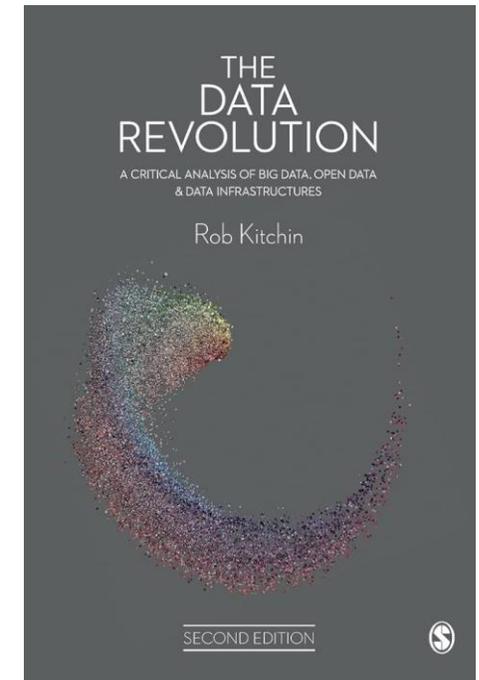
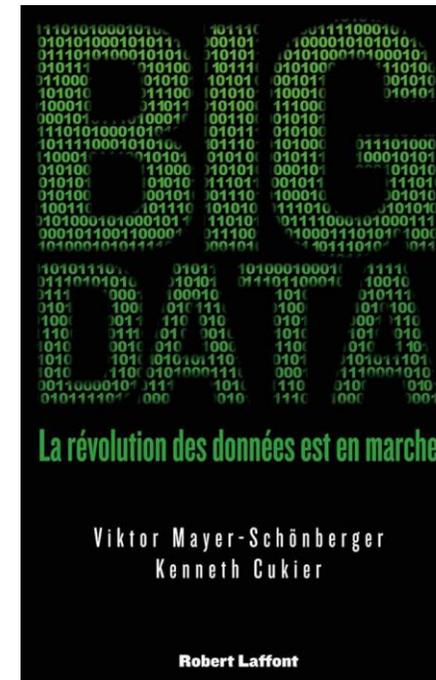
# Fascination < Réalité > Innovation

- La réalité est (encore) en deçà des promesses, mais de nombreuses innovations
  - Un processus en train de se faire avec des évolutions constantes
- Comment se positionner dans un contexte où la **fascination** l'emporte souvent **l'innovation** réelle?
  - Repenser le rôle et la place des données, les repositionner comme objet central dans la réflexion
  - Le géomaticien expert de la données



# Les données urbaines comme objet de recherche

- Avec le passage de l'informatique au numérique comme changement sociétal, les données sont le moteur des changements
- Elles sont au centre de la transformation numérique de la société comme des organisations
- Elles guident les évolutions
  - D'usages
  - Techniques
  - Méthodologiques
  - ...notamment en géomatique

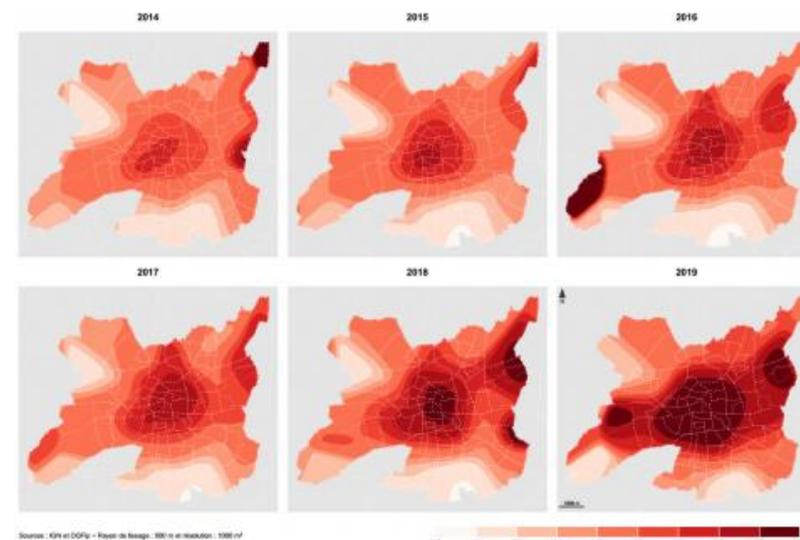
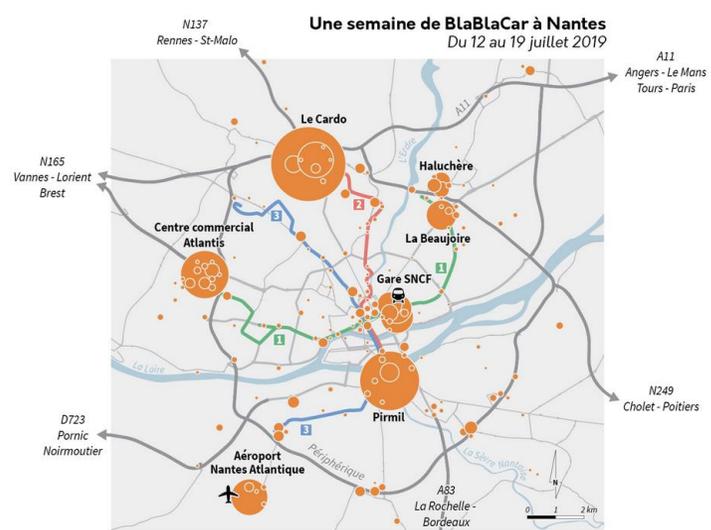
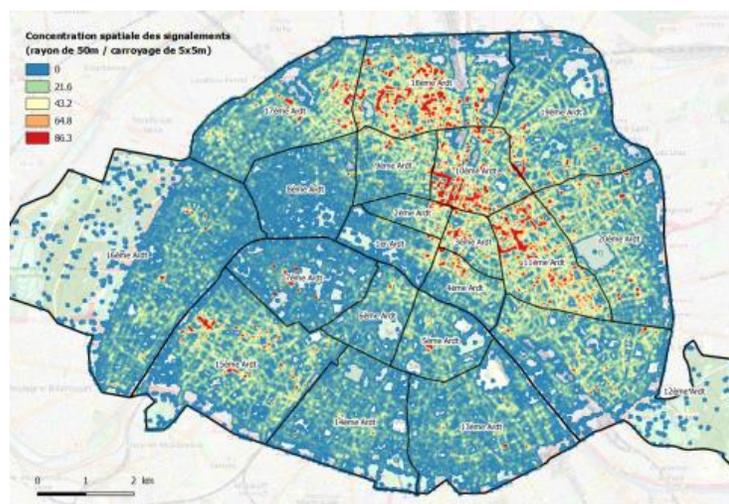


# Des approches guidées par les données

- Un pilotage par les données – Data-driven
  - Orienter et prendre des décisions stratégiques sur la base de l'analyse et de l'interprétation des données pour améliorer ses performances
  - Nouveau paradigme interprétatif, l'idée n'est plus seulement de constater, qualifier, quantifier et interpréter, mais d'anticiper et de prédire en apprenant des données
- Des domaines d'application toujours plus nombreux
  - Data-driven marketing
  - Data-driven healthcare
  - Data-driven Sustainability
  - Data-driven society
  - Data-driven science
  - Data-driven smart cities
  - Data-driven mobility
  - Data-driven urbanism
  - Data-driven geography
  - ...

# Une recherche guidée par les données

- D'abord les données ensuite la méthode !
  - Questionner et explorer les potentialités analytiques et les verrous d'usages de nouvelles bases de données dans la compréhension des dynamiques urbaines selon une approche inductive
  - Mener un travail réflexif sur la place et l'utilisabilité des nouvelles données urbaines dans des contextes de recherche ET des explorations méthodologiques



# Le mythe de la donnée



# Au départ de la réflexion

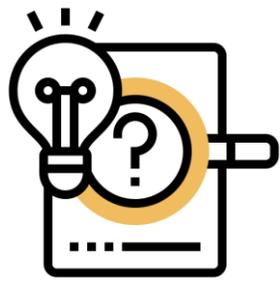
- Les données ne sont pas « données », elles ne viennent pas toutes seules, elles sont le résultats de ce qu'on a choisi de prendre, de capter, d'enregistrer, ce sont des « obtenues ».
- Les données ne sont pas « brutes », elles sont sélectionnées et *a minima* préparées.
- Les données ne sont pas neutres ni objectives, elles sont construites par des acteurs dans des contextes spécifiques et incarnent un discours.
- Les données sont normatives, tant sur ce qui est mesuré, que sur les effets qu'elles sont censées produire .
- Les données sont réductionnistes, elles mesurent certains phénomènes au détriment d'autres.
- Les données ne sont pas des entités informationnelles stables et immuables, elles sont évolutives, hétérogènes et renvoient à des caractéristiques propres.

**Agencements sociotechniques qui s'inscrivent dans des réseaux  
d'acteurs, des objectifs, des politiques et des technologies**

# Aller plus loin

- Les données ne sont pas la réalité, ce ne sont que des descriptions des modélisations numériques.
- Les données ne sont pas parfaites, elles possèdent de nombreux biais.
- Les données sont des objets techniques de plus en plus complexes.
- Les données ne sont pas prêtes à l'emploi, elles nécessitent des opérations de traitements et d'analyse pour les transformer en informations et en connaissances

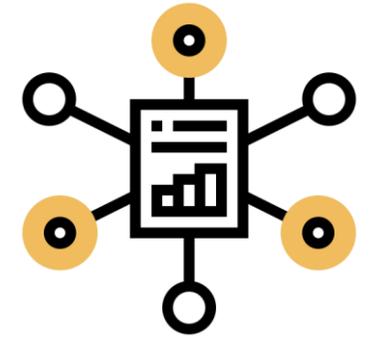
**Ressources informationnelles stratégiques pour l'analyse urbaine qui nécessitent toujours plus d'expertises méthodologiques et techniques**



# Position de recherche

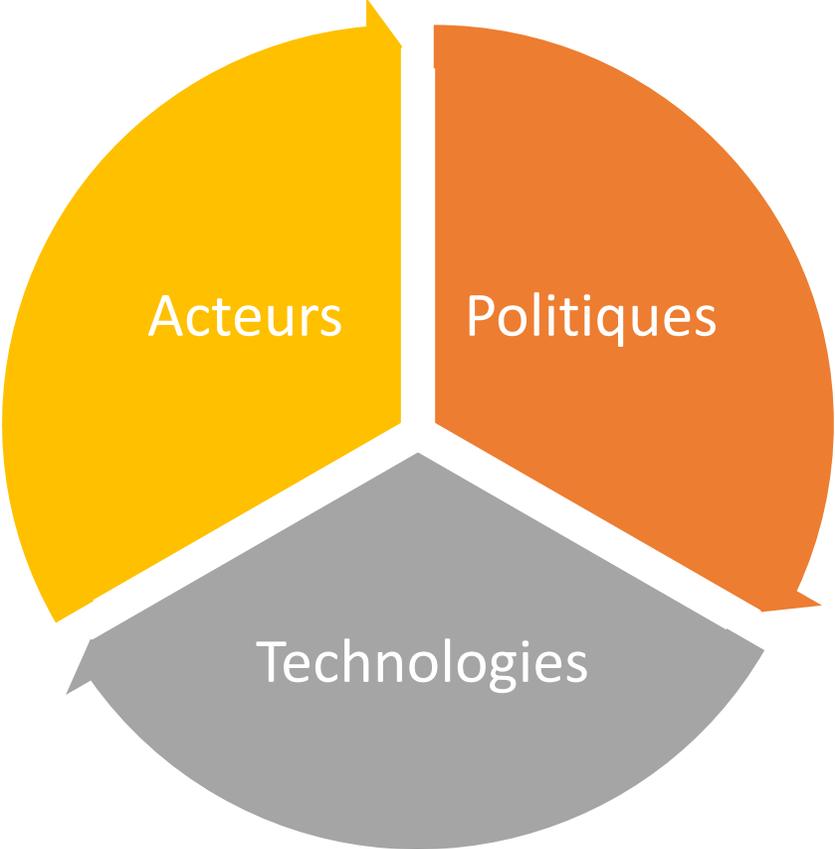
- Développer une réflexion sur l'évolution de la géomatique par le prisme des **données urbaines** comme **objet de science** et **champ d'expertises** dans une perspective critique, opérationnelle et pédagogique
- Mener une réflexion prospective autour des enjeux de la géomatique à l'heure de la mise en données des villes nécessite une double approche complémentaire :
  - Explorer et analyser les processus de production, de mise en circulation et d'utilisation des données urbaines
  - Formaliser les caractéristiques des données urbaines, investiguer leurs potentialités analytiques et réfléchir à la chaîne de traitement

# Un ensemble de cadres théoriques



- Objet frontière / Objet intermédiaire - ANT (Latour, Callon, Vinck)
  - Dispositif sociotechnique – STS (Akrich, Flichy)
- 
- Assemblage sociotechnique – CDS (Dalton, Thatcher, Kitchin)
  - Sociologie des données (Denis, Goeta)
  - Sociologie de la quantification (Desrosières)
  - Sociopolitique des données (Courmont, Cattaruzza)
  - Médiations des données urbaines (Dymytrova, Paquienséguy, Le Corf)
  - Culture des données, littéracie des données (Wolff, Wise)
- 
- SIG critique (Burns, Thatcher, Wilson, Noucher)
  - SIG et données massives (Goodchild, Lee, Kang, Thatcher)
  - Science des données spatiales ~critique~ (Anselin, Singleton, Arribas-Bel, Brunsdon)

# Triptyque sociotechnique des données urbaine



# Données = assemblages (socio + technique)

## Contextes de production et d'utilisation > Mise en données des villes, fabrique des données urbaines...

- Acteurs
- Objectifs
- Coûts, valeurs économiques
- Politiques publiques
- Modes de gouvernance
- Usages, et usagers
- Mise en circulation
- Standardisation

## Modalités d'utilisation > Outils et méthodes, la chaîne de traitement

- Licences, formats
- Structuration, modélisation spatiale, sémantique
- Granularité, complétude, précision, qualité
- Méthodes de préparation, d'analyse et de représentations
- Outils (BD, SGBD, SIG, statistiques, simulation....)

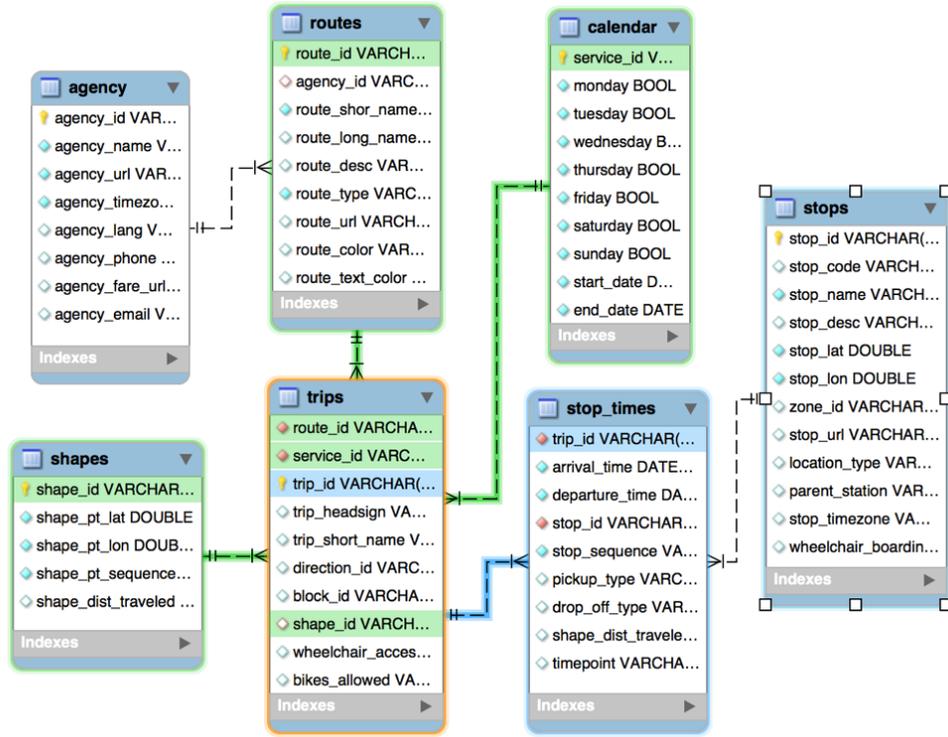
# Continuum sociotechnique des données urbaines



Remettre de la technique dans la réflexion (formats, outils, méthodes...)  
Mettre à contribution les compétences du géomaticiens !

# LOI MOBILITÉS

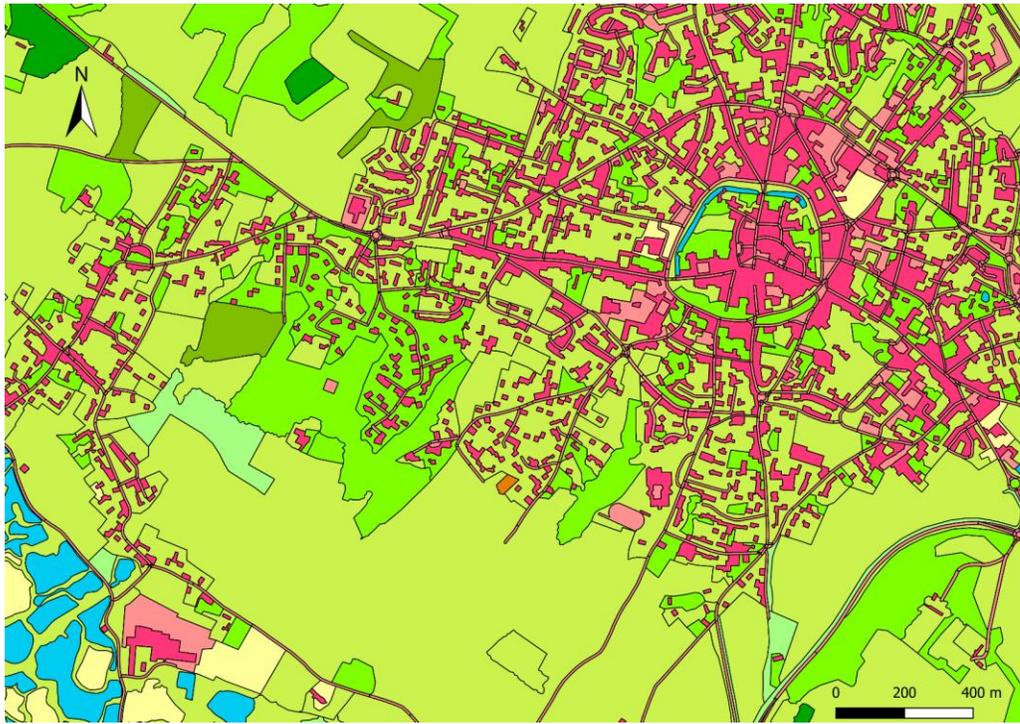
## Loi climat et résilience



From: Gare St Jean, Bordeaux  
 To: Maison des Suds, 12 Esp. des Antilles, 3

Time	Duration
09:09 – 09:49	40 min
09:16 – 09:56	40 min
09:13 – 09:56	43 min
09:13 – 10:01	48 min
09:09 – 10:01	52 min
09:33 – 10:13	40 min





CS1. Sans végétation			CS2. Avec végétation						
CS1.1 Surfaces anthropisées		CS1.2 Surfaces naturelles		CS2.1 Végétation ligneuse		CS2.2 Végétation non ligneuse			
CS1.1.1 Zones imperméables	CS1.1.2 Zones perméables	CS1.2.1 Sols nus (sable, pierres meubles, rochers saillants...)	CS1.2.2 Surfaces d'eau (continentale et maritime)	CS1.2.3 Névés et glaciers	CS2.1.1 Formations arborées	CS2.1.2 Formations arbustives et sous-arbusteuses (landes basses, formations arbustives, formations arbustives organisées...)	CS2.1.3 Autres formations ligneuses (vignes et autres lianes)	CS2.2.1 Formations herbacées (pelouses et prairies, terres arables, roselières...)	CS2.2.2 Autres formations non ligneuses (lichen, mousse, bananiers...)
CS1.1.1.1 Zones bâties		CS1.1.1.2 Zones non bâties (fruits, placis, parking...)			CS2.1.1.1 Peuplement de feuillus				
CS1.1.2.1 Zones à matériaux minéraux (pierre brute - voies ferrées, plates bordières, chemins empierrés, chantiers, carrières, salines...)		CS1.1.2.2 Zones à autres matériaux (compactés - déchargés)			CS2.1.1.2 Peuplement de conifères				
					CS2.1.1.3 Peuplement mixte				

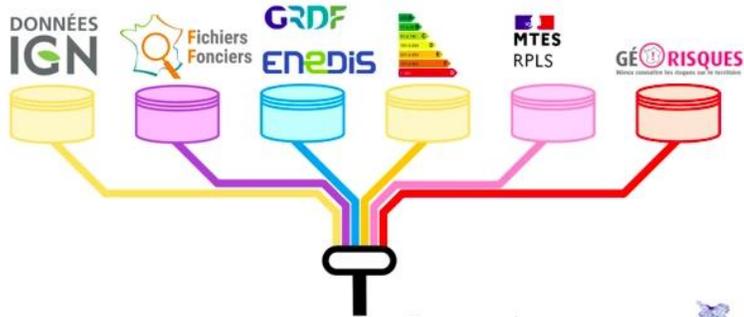
IGN

 Cerema

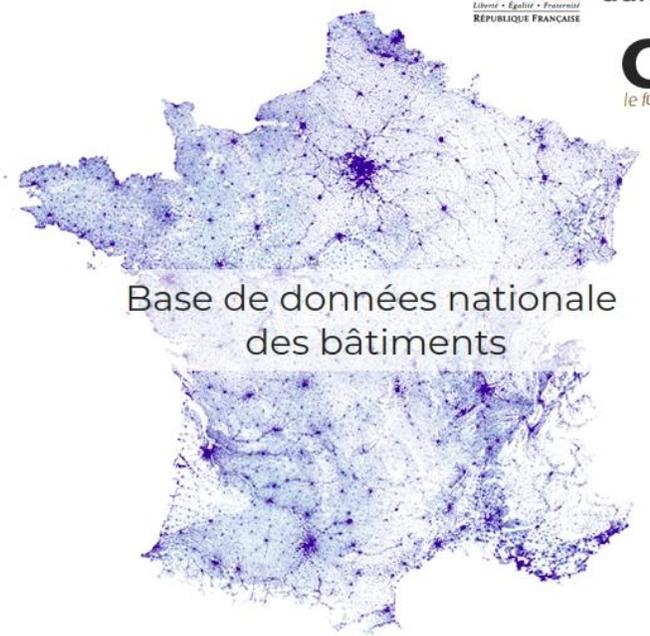
  
 MINISTÈRE  
 DE LA TRANSITION  
 ÉCOLOGIQUE  
 ET DE LA COHÉSION  
 DES TERRITOIRES  
 Liberté  
 Égalité  
 Fraternité

Conseil national de l'information géolocalisée





data.gouv.fr  
**CSTB**  
 le futur en construction

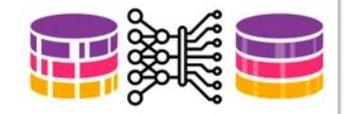


Base de données nationale des bâtiments

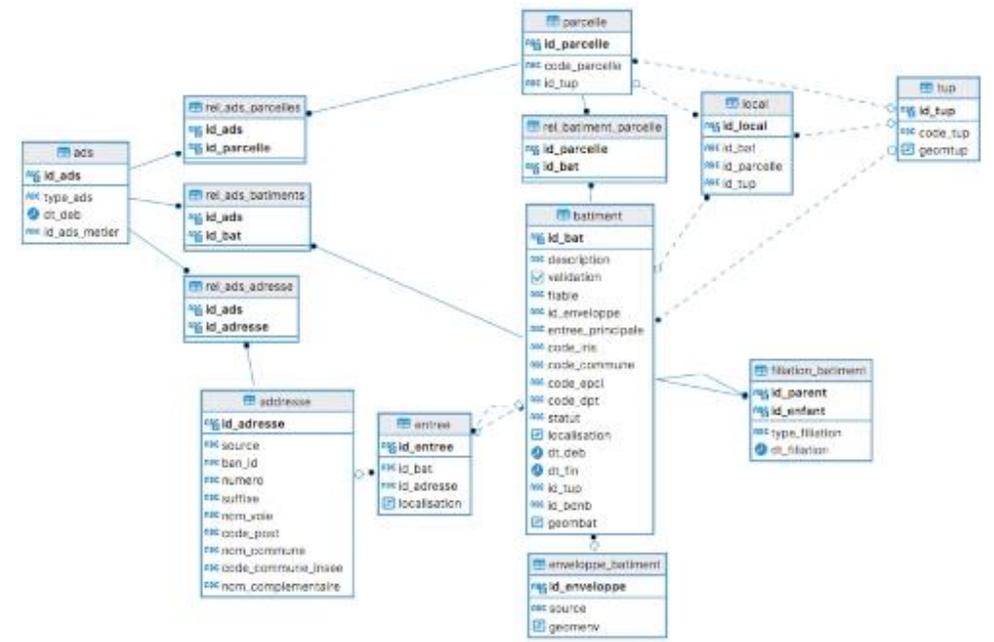
Croisement Géospatial



Prédiction valeurs manquantes



Calcul d'indicateurs métiers



DPE réels  
 BDNB open  
 (20% des logements)

DPE resimulés  
 BDNB expert  
 (100% des logements)



**API OPEN +**

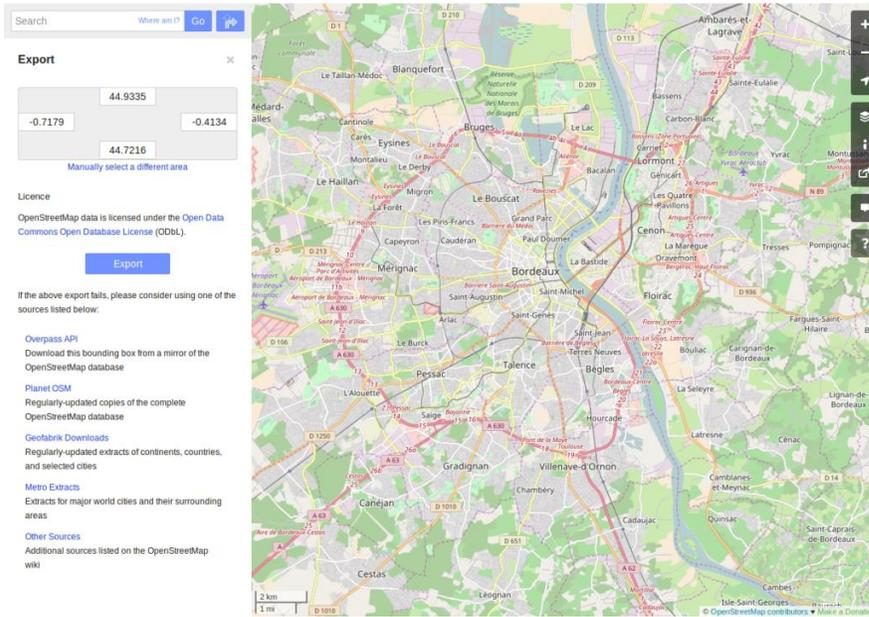
- API AVEC JETON D'AUTHENTIFICATION
- QUOTAS RÉHAUSSÉS \*
- HAUT NIVEAU DE SERVICE
- ASSISTANCE

**API EXPERT**

- DONNÉES SUPPLÉMENTAIRES ENRICHIES
- PRÉDICTIONS DES DONNÉES DPE SANS DIAGNOSTIC RÉEL
- ESTIMATION DE L'INFLUENCE DE LA VALEUR ÉNERGÉTIQUE SUR LES COÛTS DE VENTE

**API AYANTS DROITS**

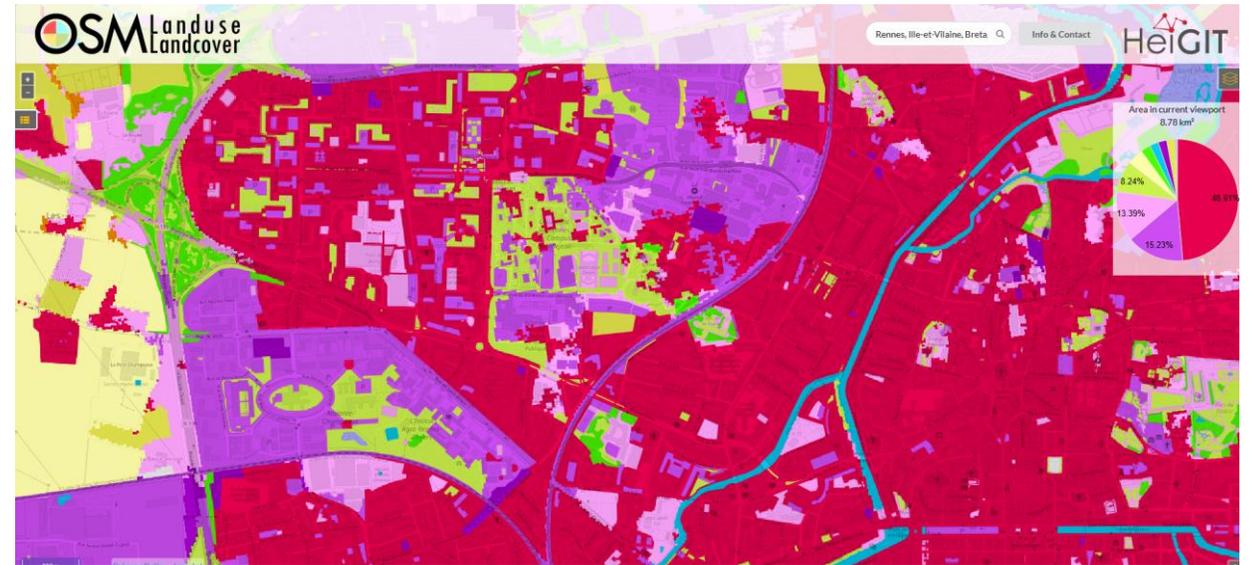
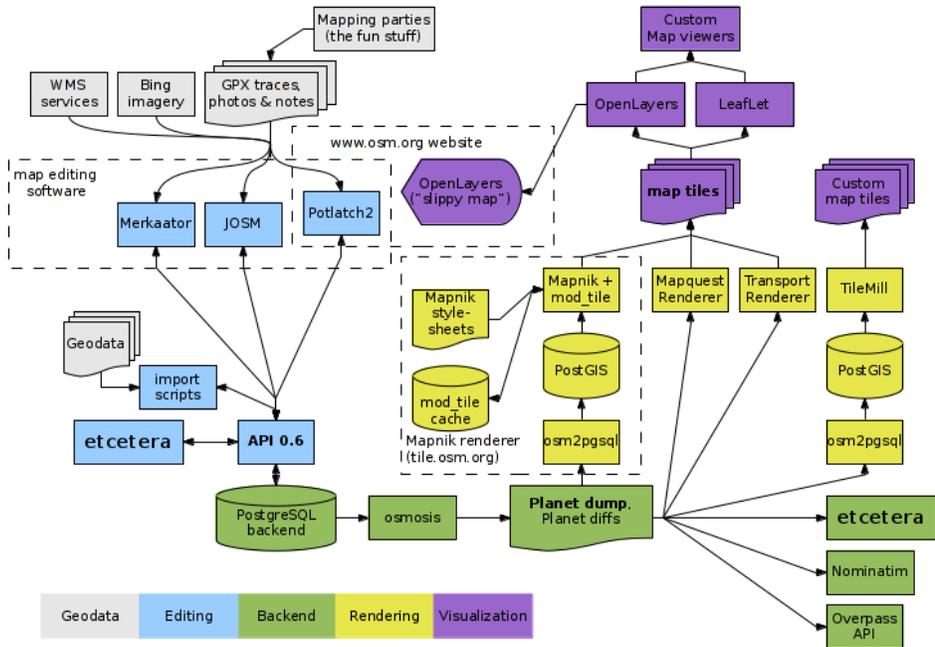
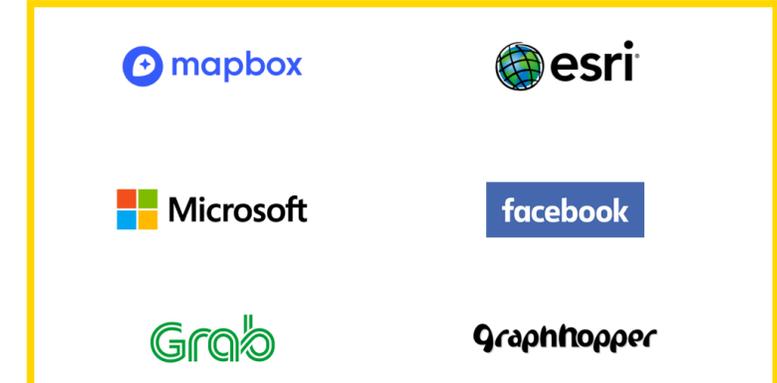
- LOTS DE DONNÉES SOUS CONVENTION
- ACCÈS AUX FICHIERS FONCIERS



Platinum Corporate Members



Gold Corporate Members



# Une multitude de données urbaines

Référentielles



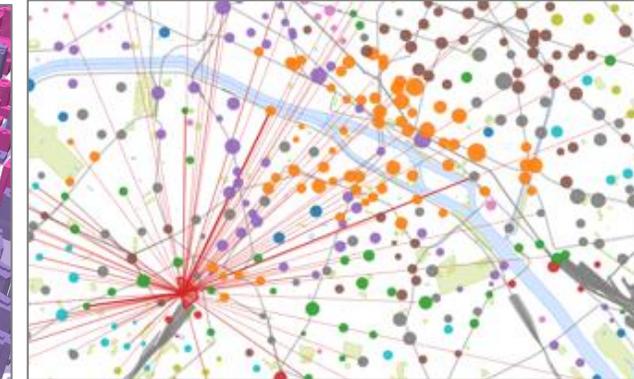
Métiers



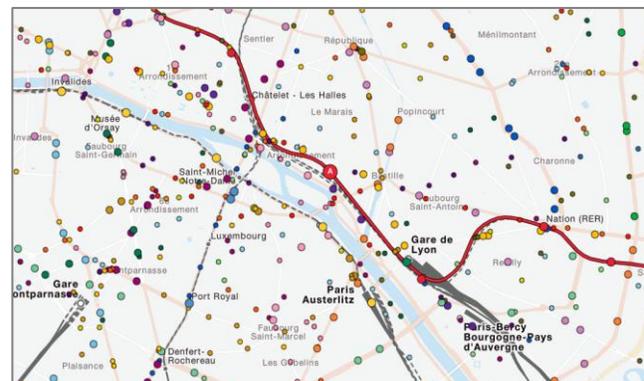
3D



Enregistrements



Temps réel



Capteurs



Contributives

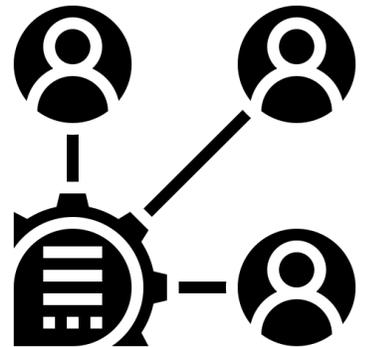


Plateformes numériques



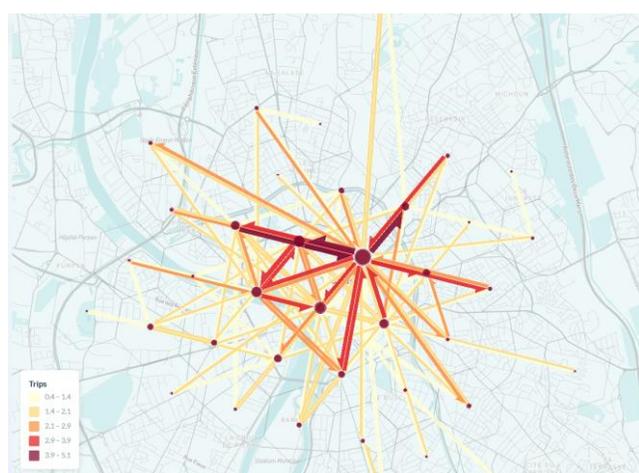
# Nouvelles dynamiques autour des données

- **Démonopolisation / Dérégulation** dans la production et la gouvernance des données urbaines
- Les données urbaines ne sont plus l'apanage des pouvoirs publics qui voient leur monopole et leur expertise remis en question
- Diversification et pluralisation des acteurs impliqués
- Apparition et consolidation de nouveaux entrants



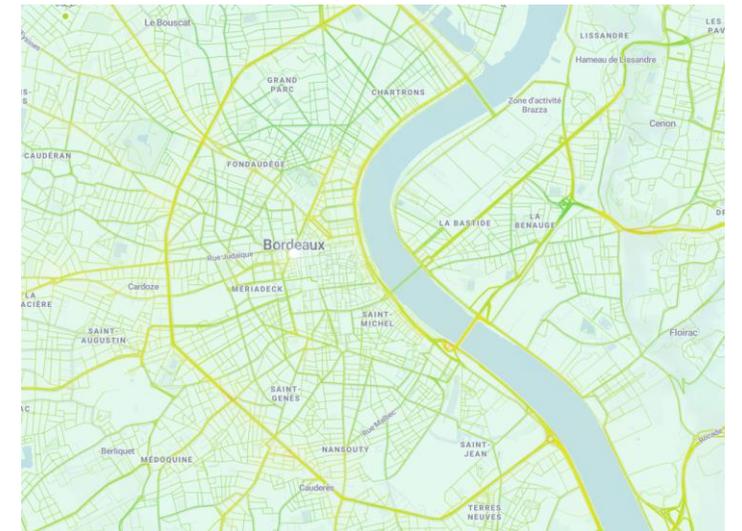
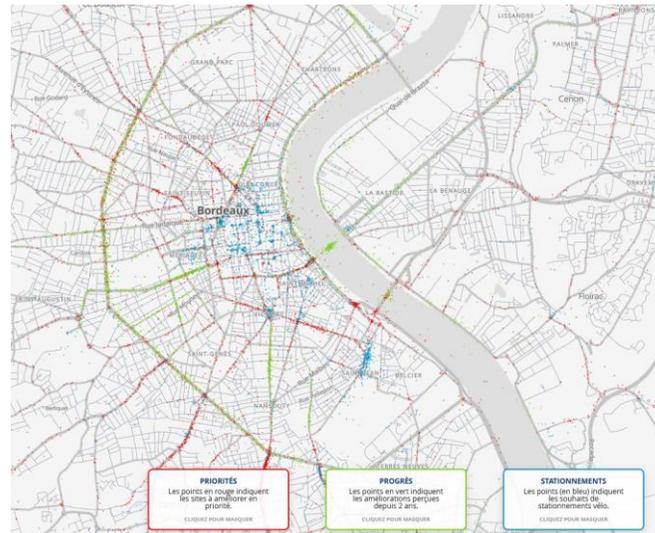
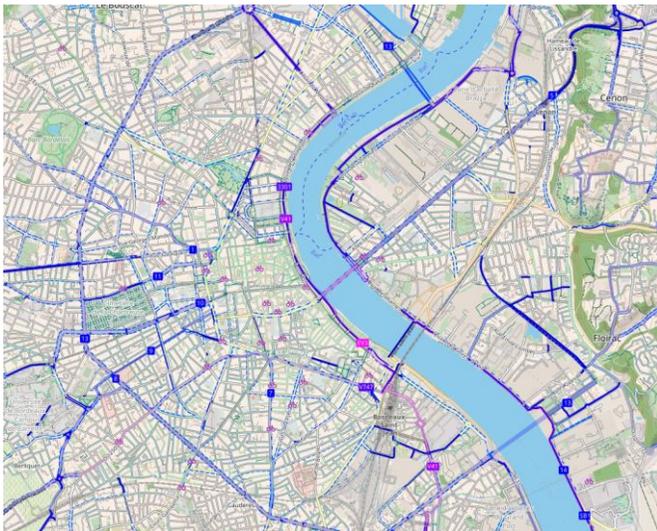
# Acteurs privés

- Bureaux d'études, DSP, opérateurs de réseaux...
  - GTFS, GBFS, Billettique, log Wifi, consommation énergétique
- Entreprises privées
  - FMD, FCD, données GPS, traces de parcours, végétation....
- Plateformes numériques
  - Waze, Uber, Strava, BlaBlaCar, Airbnb...



# Acteurs associatifs

- OpenStreetMap
  - Bâtiments, OCS, végétation, commerces, services, équipement cyclables...
- Associations autour du vélo
  - Equipements cyclables, traces GPS, signalements...
- Sciences citoyennes
  - Qualité de l'air, bruit, végétation en ville...

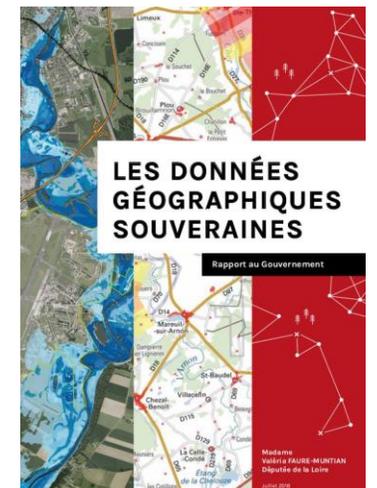


# Repositionnement des acteurs publics

- Prise de conscience depuis plusieurs années de l'importance des données dans l'action publique
- L'Etat comme chef d'orchestre des politiques des données publiques
  - Un ensemble de loi pour légiférer, réguler, reprendre le contrôle
  - Un ensemble de dispositifs pour centraliser, ouvrir, diffuser et standardiser
  - Un Etat plateforme
- Repositionnement comme garant des données
- Des initiatives locales autour des données urbaines

## Service public de la donnée : des données sur lesquelles vous pouvez compter

Le service public de la donnée créé par [l'Article 14 de la loi pour une République numérique](#) vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social. Il s'adresse principalement aux entreprises et aux administrations pour qui la disponibilité d'une donnée de qualité est critique. Les producteurs et les diffuseurs prennent des engagements auprès de ces utilisateurs. Le département Etalab de la Direction interministérielle du numérique est chargé de la mise en oeuvre et de la gouvernance de ce service public. Il référence l'ensemble des données concernées sur cette page.



# De nouvelles politiques publiques

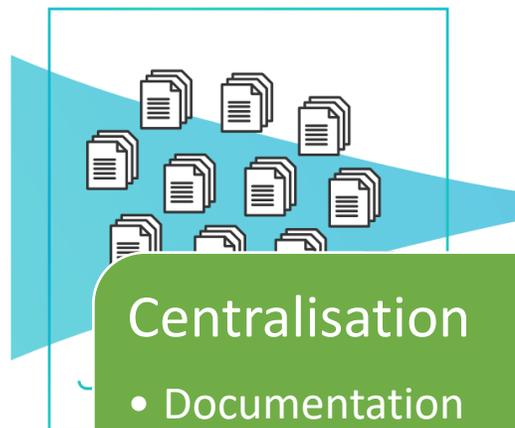


## Mise en circulation

- Open data (par défaut)
- Plateformes locales
- Licences ouvertes

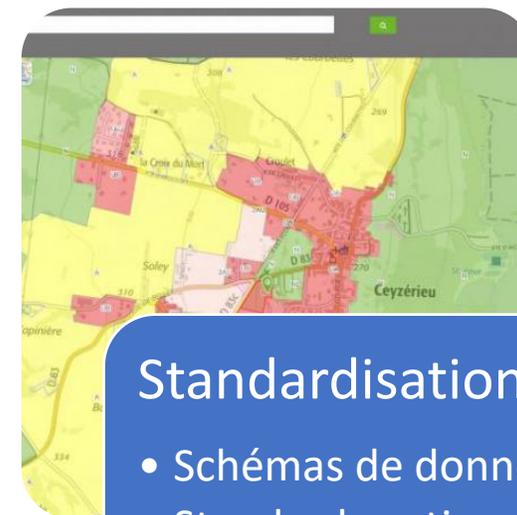


## DATA CATALOG



## Centralisation

- Documentation
- Moissonnage
- Plateformes nationales

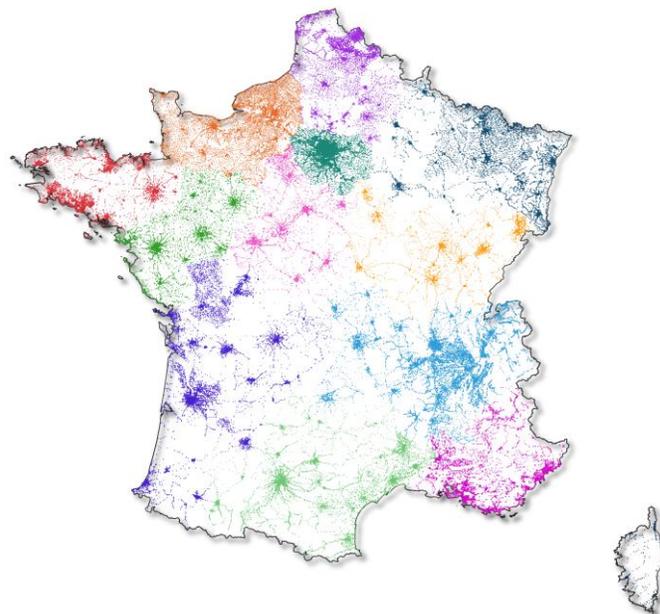
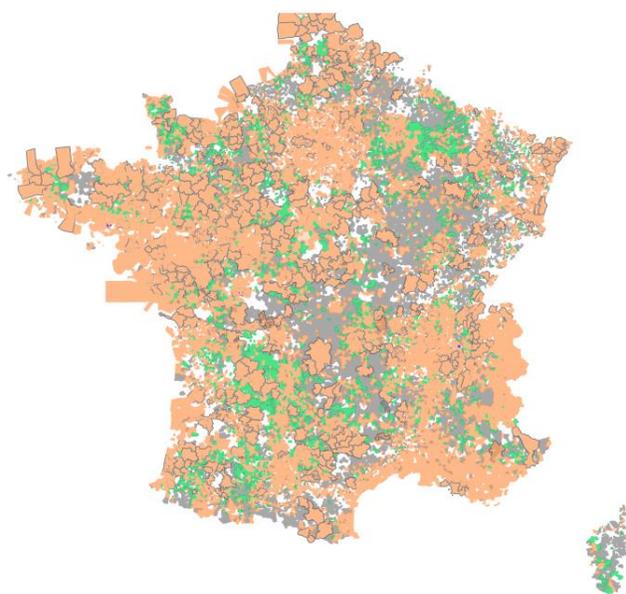


## Standardisation

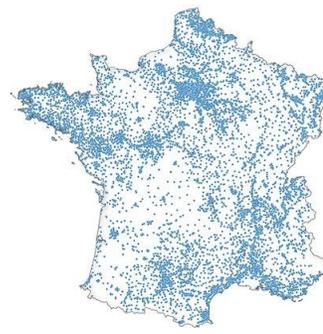
- Schémas de données
- Standards nationaux
- Bases de données nationales

# Passage à l'échelle des données urbaines

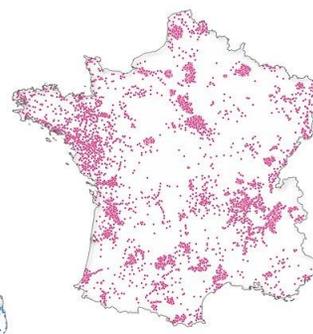
- Développer des schémas de données (Schema.data.gouv.fr)
- Multiplier les (géo)standards (CNIG, GTFS, NETEX)
- Alimenter les référentiels nationaux



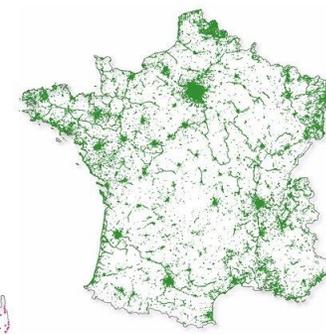
Répertoire national des  
infrastructures de recharge  
pour véhicules électriques  
57 458 points



Base nationale des  
lieux de covoiturage  
9184 points



Base nationale des  
aménagement cyclables  
219 143 tronçons



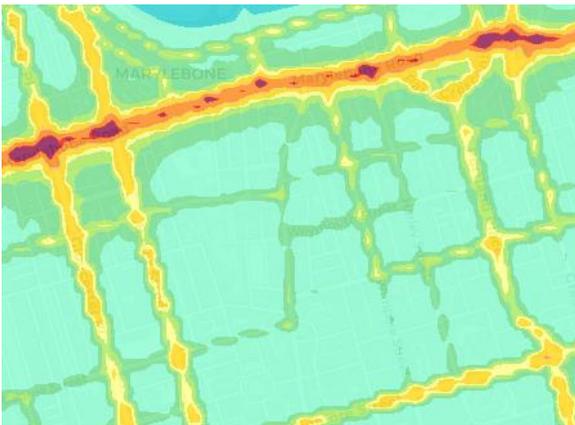
# Ouvrir le capot des données mettre les mains dans le cambouis

- La multitude d'informations toujours plus nombreuses et précises sont porteuses de nombreuses promesses pour l'analyse urbaine
  - Elles soulèvent des **interrogations** (nature des données, modalités de mise en circulation, outils et méthodes de traitement associés...)
- La pléthore de nouveaux gisements informationnels disponibles ne sont pas des filons d'informations facilement accessibles, prêts à l'emploi et aisément exploitables
  - La production de connaissances à partir des données nécessite un temps de découverte et de compréhension, l'utilisation de méthodes et le recours à des outils toujours plus complexes



# Type de données

## Raster



## Vecteur



## Tableaux

Prévisualiser							
Structure des données				Métadonnées			
id_zone	INSEE_C	LIBGEO	EPCI	DEP	REG	loypredm2	lwrIpM2
1	67462	Sélestat	246700967	67	44	102495505031762	808630416530593
1	67461	Schwobsheim	200030526	67	44	980242068322335	773007909082588
1	67360	Ohnheim	200030526	67	44	980242068322335	773007909082588
1	67239	Kintzheim	246700967	67	44	980242068322335	773007909082588
1	67056	Bootzheim	200030526	67	44	980242068322335	773007909082588

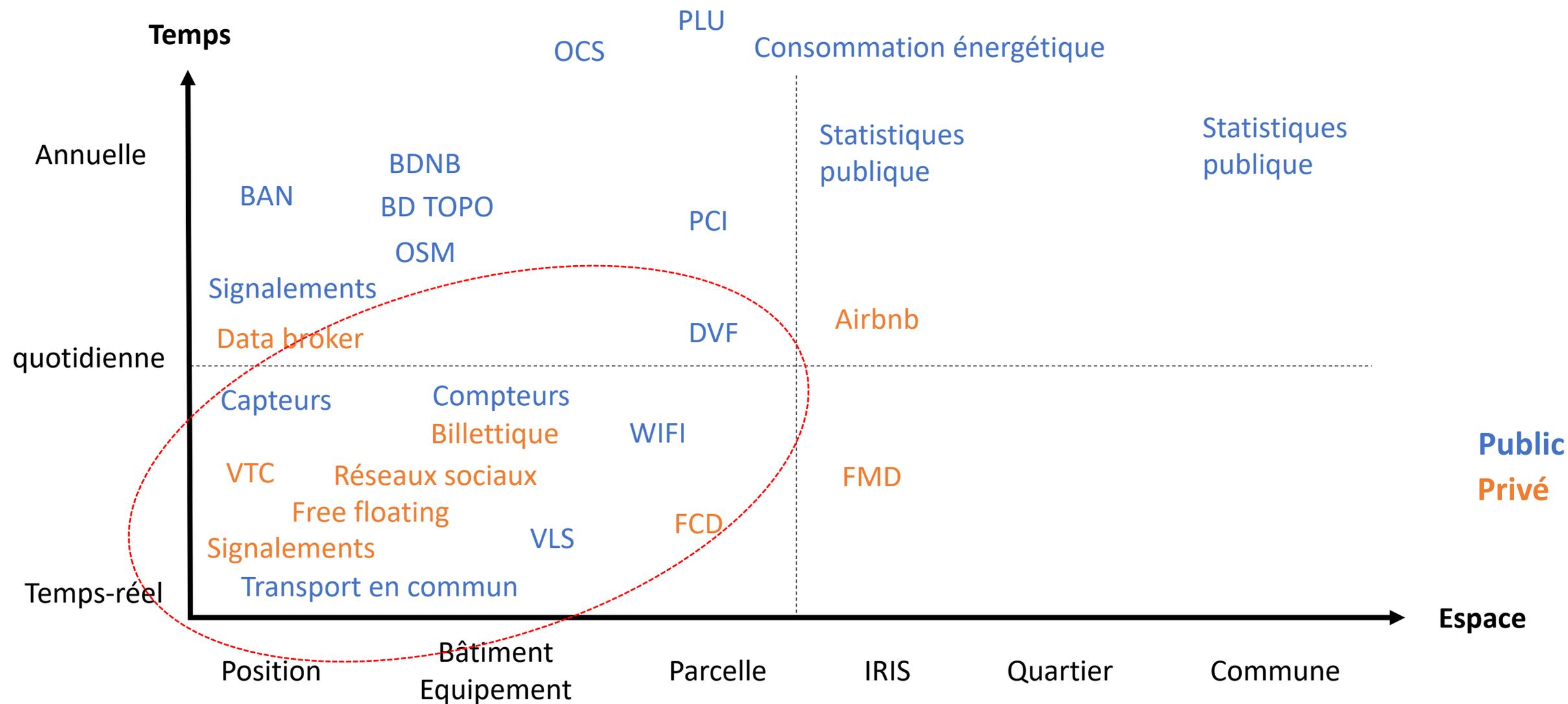
Code du quartier prioritaire	Nom du quartier prioritaire	Commune	Code INSEE	EPCI	Départem
QP078016	Vieux Pays	Sartrouville	78586	CA Saint-Germain Boucles de Seine	YVELINES
QP049001	Belle Beille	Angers	49007	CU Angers Loire Métropole	MAINE-ET
QP011004	Bastide Pont-Vieux	Carcassonne	11069	CA Carcassonne Agglo	AUDE
QP089006	Arènes - Champs Plaisants	Sens	89387	CA du Grand Sénonais	YONNE
QP030003	Pissevin - Valdegour	Nîmes	30189	CA de Nîmes Métropole	GARD
QP011010	Centre Ville	Lézignan-Corbières	11203	CC Région Lézignanaise, Corbières e...	AUDE
QP078007	Jean Macé	Trappes	78621	CA de Saint-Quentin-en-Yvelines	YVELINES
QP059072	Pacot - Vandracq	Lambersart	59328	Métropole Européenne de Lille	NORD
QP024001	Rive Gauche	Bergerac	24037	CA Bergeracoise	DORDOGN
QP040002	Cuyès	Dax	40088	CA du Grand Dax	LANDES
QP042006	Montcel - Centre Ville	La Ricamarie	42183	Saint-Etienne Métropole	LOIRE
QP075019	Les Portes Du Vingtième	Paris 20ème arrondissement	75056	Métropole du Grand Paris	PARIS
QP038008	Teisseire Abbaye Jouhaux Châtelet	Grenoble	38185	Métropole Grenoble-Alpes-Métropole	ISERE
QP057019	Hauts De Vallières	Metz	57463	Metz Métropole	MOSELLE
QP059043	Coeur D'Etoile	Le Quesnoy	59481	CC du Pays de Mormal	NORD

# Données structurées / Semi-structurées

bike_id	current_range_meters	is_disabled	is_reserved	last_reported	lat	lon	pricing_plan_id	rental_uris.android
d352959f-cb5b-4cdc-9e3c-7248348d6994	16500	false	false	1700059893	45.22742	5.670037	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d3694478-2ee0-46e4-8ea1-f9144537c314	15900	false	false	1700060495	45.19691	5.776585	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d36bcb14-a5e3-4c6f-ae89-c82192a02894	300	false	false	1699737386	45.18774	5.714836	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d3961d9a-f316-4f28-9905-a5f25f8c6370	18300	false	false	1700058862	45.15698	5.735299	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d3bdaf6a-6830-4bb4-a9c4-9c2c2ccf4c14	25500	false	false	1700060041	45.20458	5.663698	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d3b5c80f-2bb6-45b8-bda4-548962e9e788	26400	false	false	1700060025	45.22063	5.797415	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d03187c1-f583-46ea-92e8-3146bd94d100	4500	false	false	1700059646	45.23729	5.660563	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d043a317-570e-4388-a6f5-1ae3a052ab90	14400	false	false	1700060608	45.16846	5.730155	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d0748d21-4e58-4f62-ac85-e5c835ee7670	38500	false	false	1700060580	45.21085	5.665482	3cdb04f7-bd49-58b7-9ce8-8a38b79d8f50	https://go.ridedott.com/vehicl
d07cb0a5-ab68-488e-97d5-dd52d0fc1f88	7800	false	false	1700059841	45.15862	5.682165	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d08d1444-a243-442f-8ad6-d81cd20ab3d4	14500	false	false	1700060608	45.23795	5.688584	3cdb04f7-bd49-58b7-9ce8-8a38b79d8f50	https://go.ridedott.com/vehicl
d099721a-1b24-4be2-836a-9cc69a382550	44000	false	false	1700059901	45.19062	5.775954	3cdb04f7-bd49-58b7-9ce8-8a38b79d8f50	https://go.ridedott.com/vehicl
d0a123a4-6f3b-4aed-a750-697517b6f3c0	12600	false	false	1699910777	45.14789	5.693052	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d0c515e8-c6d8-42fb-a8bd-e761f0a46ff0	39000	false	false	1700060322	45.18498	5.785346	3cdb04f7-bd49-58b7-9ce8-8a38b79d8f50	https://go.ridedott.com/vehicl
d0c8d6b9-4225-42b8-ada6-1f2745799c20	22800	false	false	1700059764	45.21785	5.779563	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d12b7bb0-e178-46e4-ad5d-e688e2659750	3500	false	false	1700060564	45.21983	5.678540	3cdb04f7-bd49-58b7-9ce8-8a38b79d8f50	https://go.ridedott.com/vehicl
d13628f3-67e3-4426-8dd3-6b4094e2ddd8	300	false	false	1700046134	45.16745	5.712297	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d15dc4bc-8601-424e-ad91-c145b0ca2260	26400	false	false	1700060285	45.17549	5.720653	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d15ec886-82f2-4e36-bb6f-12d1f39c9020	20700	false	false	1700060631	45.13955	5.720219	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d1642b3e-9a22-4a29-8a1d-2aa4407b5f40	26400	false	false	1700060190	45.14798	5.692920	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d181520c-951c-4e8d-8f70-4d09533e2ce8	9300	false	false	1700060232	45.21364	5.652583	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d184e4c5-b8b2-46b1-9e27-38e9a373fc78	24300	false	false	1700060462	45.21113	5.750765	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d1a2a8ef-d2cb-4235-b75b-65a57319d6b4	14400	false	false	1700059864	45.22634	5.802040	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d1be6086-bad6-4cc8-9d47-c272bb3d4cd8	13500	false	false	1700060215	45.17904	5.720427	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d1b853f4-3d46-4151-b86c-5ea354163b70	19200	false	false	1700059734	45.18336	5.726568	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl
d1ba15dd-6cee-45fc-a2bb-41b46d7e344c	24300	false	false	1700060265	45.20264	5.744505	07a0a280-4c21-4d20-bd07-0dcde83ba6e4	https://go.ridedott.com/vehicl

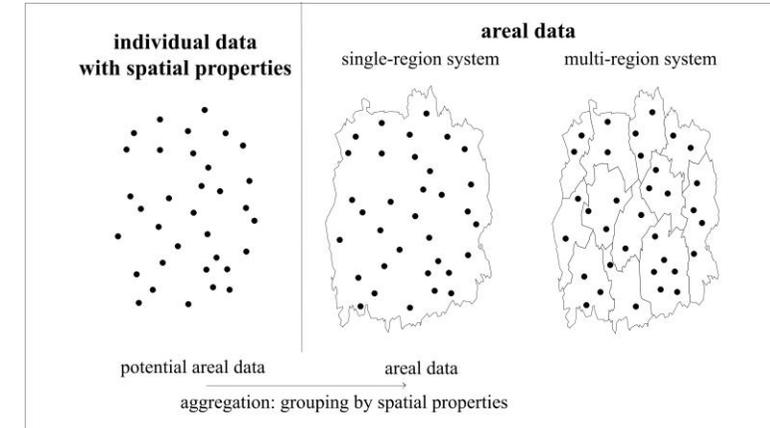
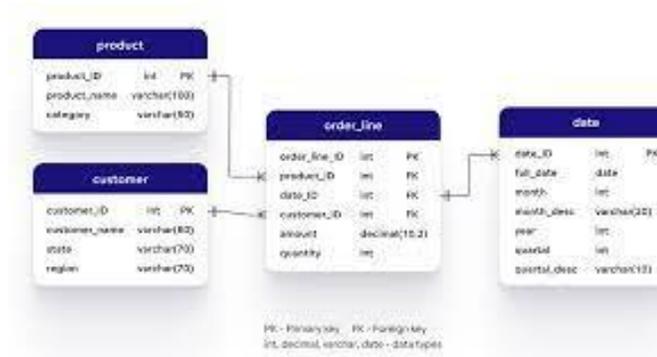
```
data:
  bikes:
    0:
      bike_id: "d352959f-cb5b-4cdc-9e3c-7248348d6994"
      current_range_meters: 16500
      is_disabled: false
      is_reserved: false
      last_reported: 1700059893
      lat: 45.227425
      lon: 5.670037
      pricing_plan_id: "07a0a280-4c21-4d20-bd07-0dcde83ba6e4"
      rental_uris:
        android: "https://go.ridedott.com/vehicles/QYK451?platform=android"
        ios: "https://go.ridedott.com/vehicles/QYK451?platform=ios"
      vehicle_type_id: "dott_scooter"
    1:
      bike_id: "d3694478-2ee0-46e4-8ea1-f9144537c314"
      current_range_meters: 15900
      is_disabled: false
      is_reserved: false
      last_reported: 1700060495
      lat: 45.196906
      lon: 5.776585
      pricing_plan_id: "07a0a280-4c21-4d20-bd07-0dcde83ba6e4"
      rental_uris:
        android: "https://go.ridedott.com/vehicles/XNWF5?platform=android"
        ios: "https://go.ridedott.com/vehicles/XNWF5?platform=ios"
      vehicle_type_id: "dott_scooter"
    2:
      bike_id: "d36bcb14-a5e3-4c6f-ae89-c82192a02894"
      current_range_meters: 300
      is_disabled: false
      is_reserved: false
      last_reported: 1699737386
      lat: 45.187737
      lon: 5.714836
      pricing_plan_id: "07a0a280-4c21-4d20-bd07-0dcde83ba6e4"
      rental_uris:
        android: "https://go.ridedott.com/vehicles/JE286J?platform=android"
        ios: "https://go.ridedott.com/vehicles/JE286J?platform=ios"
      vehicle_type_id: "dott_scooter"
```

# Résolutions spatiotemporelles des données

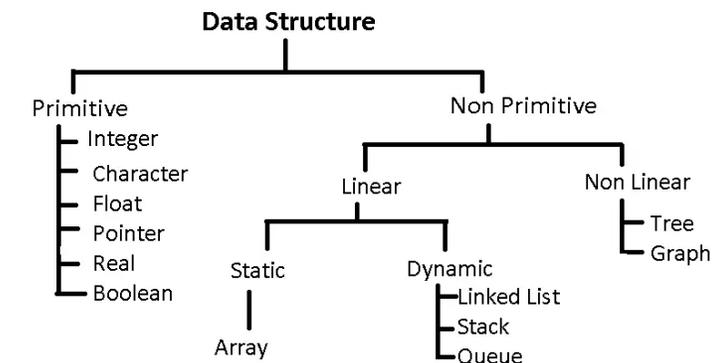


# Caractéristiques des données urbaines

- Volume
- Granularité
- Format
- Variables
- Encodage
- Structure
- Modèle de données
- Composante spatiale
- Topologie
- Dimension temporelle
- Qualité
- Utilisabilité
- Historicité
- ...



	CSV	Parquet	JSON
Read Speed	✓	✓	
Small File Size		✓	
Splittable	✓	✓	✓
Included Data Types		✓	✓
Easy to Read	✓		✓
Nestable		✓	✓
Columnar		✓	
Complex Data Structures		✓	✓





# Des données aux connaissances



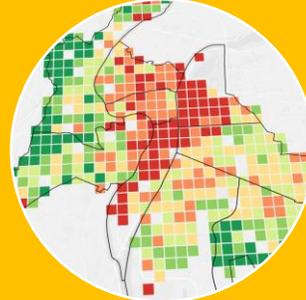
## Données

Descriptions  
élémentaires d'une  
réalité



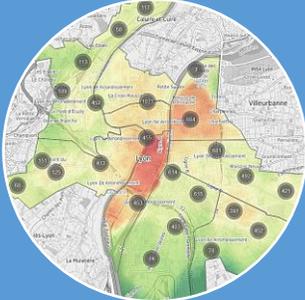
## Informations

Données  
contextualisées et  
préparées



## Connaissances

Informations  
organisées et  
analysées



## Intelligence

Connaissances  
appliquées à un  
domaine



Récupération

Préparation

Exploration

Traitement

Analyse

Modélisation

Prédiction



# Récupération des données

- Trouver les « bonnes » et les « nouvelles » données urbaines
- Accessibilité aux données Physique vs Intellectuelle
- Jeux de données vs flux (API, WMS, WFS)
- Données millésimée, données historicisées ou données en temps réel

## Console API

Consultez la [documentation de l'API v1 de recherche](#) si vous souhaitez plus d'informations.

Lister les jeux de données

Requête en texte intégral

rows: 10

start: 0

sort:

facet:

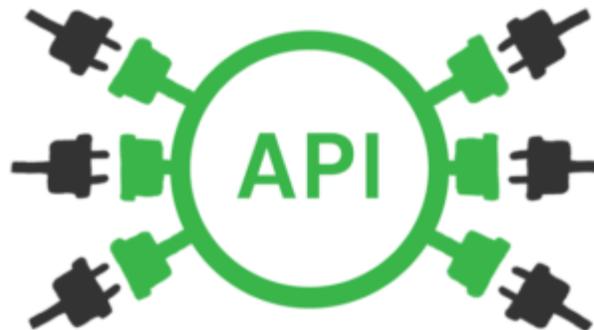
refine:

exclude:

Envoyer

GET /api/datasets/1.0/search/

```
{
  "results": 500,
  "parameters": {
    "rows": 10,
    "start": 0,
    "staged": false,
    "include_drafts": false,
    "types": "Open",
    "timezone": "UTC"
  },
  "datasets": [
    {
      "datasetid": "logpersonestacon",
      "meta": {
        "domain": "public",
        "staged": false,
        "visibility": "domain",
        "metadata_processed": "2015-08-05T13:11:23.17000+00:00",
        "data_processed": "2017-03-27T12:30:00+00:00",
        "modified": "2015-12-11T14:08:00+00:00",
        "license": "Open License v1.0",
        "publisher": "cdm"
      },
      "description": "opAPL : Allocation de Logement Familiale ALS : Allocation de Logement Social APL : Aide Personnalisée au Logement/3p-3pces trois aides ne sont pas cumulables. L'ordre de priorité d'attribution est le suivant : APL, APL, ALS-3p-3pces extraites grâce au projet ca href="https://github.com/angilles/open-moulinette">Open Moulinette/3p-3p",
      "theme": [
        "Administration, Government, Public finances, Citizenship"
      ],
      "keyword": [
        "Allocation",
        "Aide au Logement",
        "Logement",
        "APL"
      ]
    }
  ]
}
```



## Index of bucket "tripdata"

Name	Date Modified	Size	Type
201307-201402-citibike-tripdata.zip	Jan 18th 2017, 11:23:25 pm	178.26 MB	ZIP file
201307-citibike-tripdata.zip	Jan 18th 2017, 11:23:27 pm	27.07 MB	ZIP file
201308-citibike-tripdata.zip	Jan 18th 2017, 11:23:27 pm	32.09 MB	ZIP file
201309-citibike-tripdata.zip	Jan 18th 2017, 11:23:27 pm	33.16 MB	ZIP file
201310-citibike-tripdata.zip	Jan 18th 2017, 11:23:28 pm	33.07 MB	ZIP file
201311-citibike-tripdata.zip	Jan 18th 2017, 11:23:28 pm	21.62 MB	ZIP file
201312-citibike-tripdata.zip	Jan 18th 2017, 11:23:28 pm	14.31 MB	ZIP file
201401-citibike-tripdata.zip	Jan 18th 2017, 11:23:29 pm	9.70 MB	ZIP file
201402-citibike-tripdata.zip	Jan 18th 2017, 11:23:29 pm	7.25 MB	ZIP file
201403-citibike-tripdata.zip	Jan 18th 2017, 11:23:29 pm	14.13 MB	ZIP file
201404-citibike-tripdata.zip	Jan 18th 2017, 11:23:29 pm	21.41 MB	ZIP file
201405-citibike-tripdata.zip	Jan 18th 2017, 11:23:29 pm	27.59 MB	ZIP file
201406-citibike-tripdata.zip	Jan 18th 2017, 11:23:29 pm	29.90 MB	ZIP file
201407-citibike-tripdata.zip	Jan 18th 2017, 11:23:30 pm	30.89 MB	ZIP file
201408-citibike-tripdata.zip	Jan 18th 2017, 11:23:30 pm	30.63 MB	ZIP file
201409-citibike-tripdata.zip	Jan 18th 2017, 11:23:30 pm	30.25 MB	ZIP file
201410-citibike-tripdata.zip	Jan 18th 2017, 11:23:30 pm	26.15 MB	ZIP file

# Préparation des données

- Nettoyage / Apurement
- Restructuration
- Filtre
- Spatialisation
- Agrégation
- Enrichissement
- ...

ID	DEPARTMENT	PHONE NUMBER	ZIP	CITY	STATE
1	Fire Department	718-999-FDNY	10004	New York	NY
2	Community Affairs	718-999-1438	60611	Chicago	IL
3	EMS Command	718-999-2770/1783	60611	Chicago	IL
4	Human Resources	718-999-2164	90054	Los Angeles	CA
5	HR	718-999-2164	90054	LA	CA
6	Intern Program	718-999-2181		SF	CA

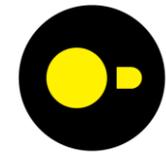
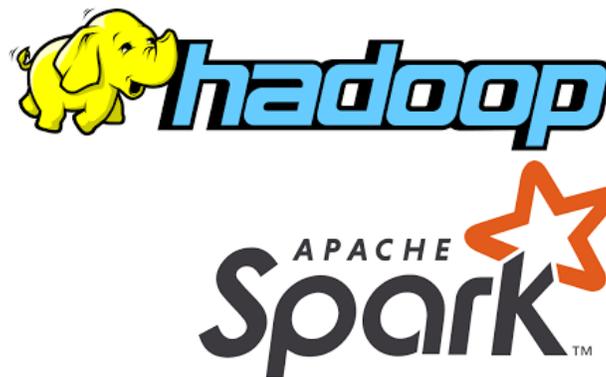
Annotations:

- Ambiguous value: points to '718-999-FDNY' in row 1.
- Formatting Rule Violation: points to '718-999-2770/1783' in row 3, with the pattern '###-###-####' below it.
- Missing Value: points to the empty 'ZIP' cell in row 6.
- Functional Dependency Violation: points to 'LA' in row 5, with the text 'ZIP -> CITY' next to it.



# Stockage et manipulation des données

- Structurer des données multi-sources
  - Archiver / Documenter
- Combiner des données hétérogènes
  - Structuration / Granularité / Temporalité
- Stocker et manipuler des données volumineuses
  - Indexer / Paralléliser



**DuckDB**





# Analyse des données

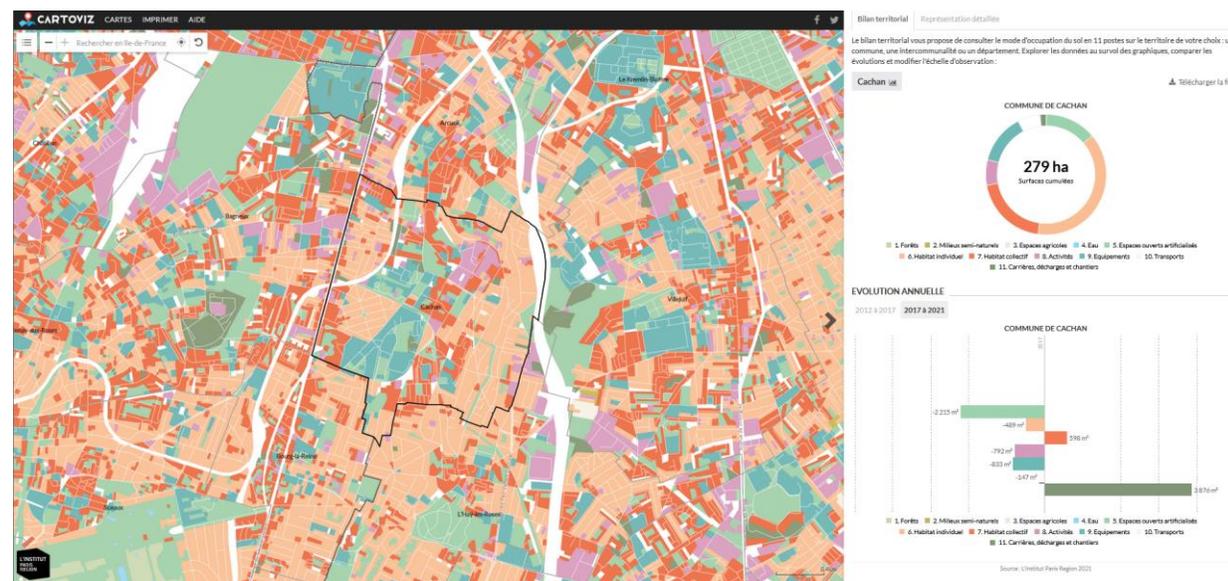
- Analyse spatiale
  - Localisation / Concentration / Distance / Agrégation / Analyse multicritères
- Analyses statistiques
  - Univarié / Multivarié (ACP, ACM, AFC, CAH, clustering)
- Analyses géostatistiques / statistiques spatiales
  - Interpolation / Autocorrélation / Lissage
- Modélisation
- Machine learning
  - Approches supervisées et non supervisées





# Valorisation des données

- Cartographie statique
  - Cartographie thématique / Print / Numérique / Communication
- Visualisation de données
  - Dataviz statique ou dynamique
- Cartographie en ligne
- Géovisualisation
- Tableaux de bord
- Data Storytelling
- Animation

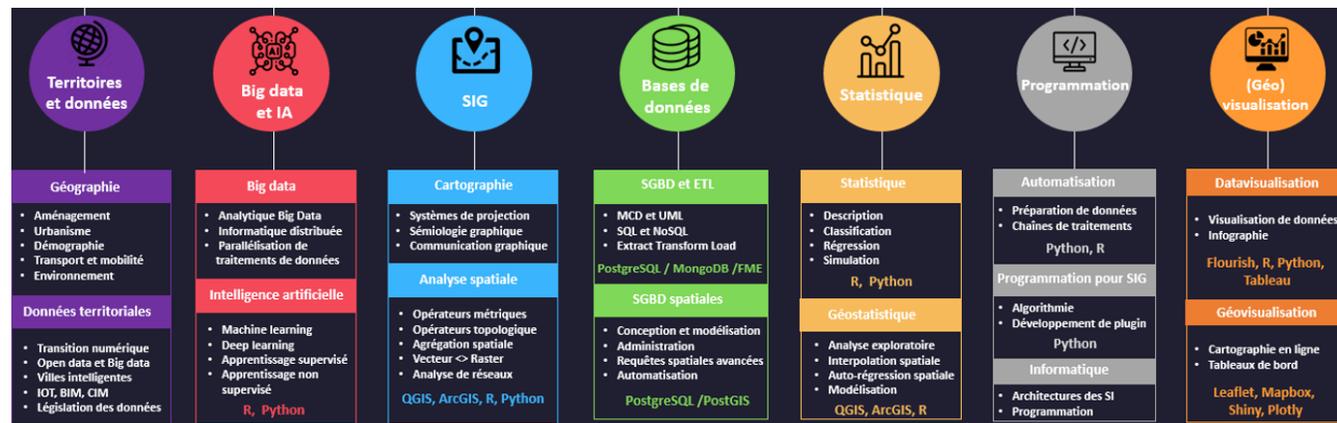


# Mise à distance

- De manière paradoxale alors que les données n'ont jamais été aussi nombreuses, on assiste à une logique de mise à distance
  - Le citoyen/usager est mis à distance des données qu'il produit lui-même au quotidien (dépossession des données personnelles)
  - Le thématicien est mis à distance des données qu'il pourrait utiliser
  - Le géomaticien est mis à distance des données qu'il utilisait
- Vers de nouveaux géomaticiens ?
  - Appropriation des outils et des méthodes par de nouveaux profils

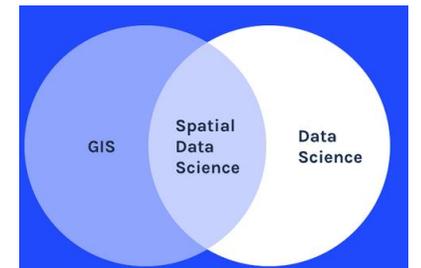
# Manque de compétences

- Les besoins du monde professionnels évoluent rapidement
  - De plus en plus de structures sont à la recherche de « moutons à cinq pattes » capables la fois de croiser, structurer, analyser et représenter de gros volumes de données géographiques, voire de proposer des approches prédictives
- Dans la pratique, les pouvoirs publics comme les entreprises privées qui manquent d'expertises internes connaissent des difficultés à recruter ces nouveaux profils aux compétences rares et recherchées



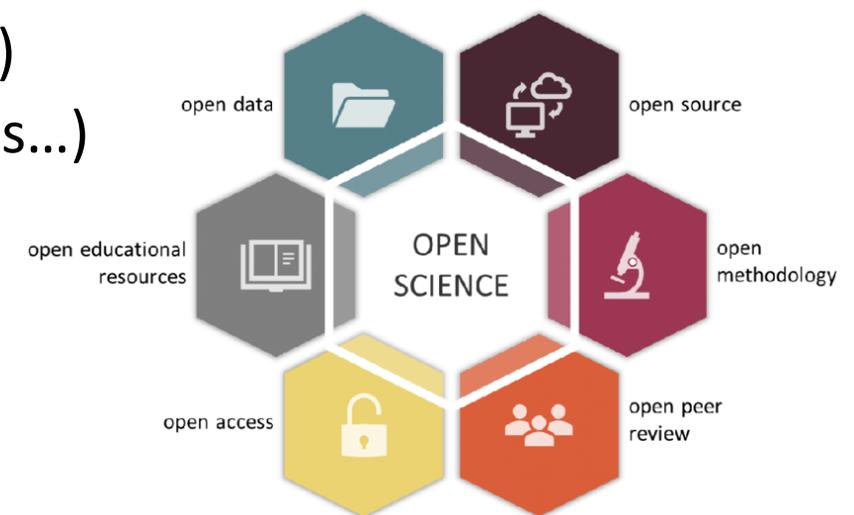
# Repenser la géomatique

- La géomatique se voit largement questionnée par le développement de la science des données tant d'un point de vue technique, méthodologique qu'épistémologique
- De nouveaux champs de compétences à investiguer
  - Sciences des données spatiales, Science des données urbaines...
  - En France, la géomatique et la science des données restent encore éloignées et peu de passerelles existent entre les communautés
  - Il existe des synergies à développer entre ces deux mondes, pour à la fois repenser la géomatique à l'heure des données massives et enrichir la science des données d'une réelle dimension spatiale



# Repenser la recherche en géomatique

- Ouvrir, ouvrir, ouvrir...
  - Utilisation de données ouvertes
  - Mobilisation et développement de solutions open source
  - Ouverture des données de la recherche
  - Partager et diffuser (dépôt de codes)
  - Documentation des méthodes (ouvrir les boites noires)
  - Reproductibilité des méthodes (scripts, notebook...)
  - Valoriser autrement (data paper, codes, applications...)



# Repenser la formation initiale et continue

- Vers un littéracie des données urbaines
  - Capacité d'identifier, de collecter, de traiter, d'analyser et d'interpréter des données afin de comprendre les phénomènes, les processus, les comportements qui les ont générées en sachant faire preuve d'esprit critique
- Former aux enjeux sociaux, économique et politiques des données
  - Gouvernance, licences, usages, éthiques...
- Former aux compétences nécessaires aux traitements des données
  - Caractéristiques des données, préparation, structuration, analyse, valorisation
  - Statistique, géomatique, dataviz, informatique, science des données...





# Perspectives

- Beaucoup de fascination qui rend ce domaine très dynamique et fertile aux recherches
- Beaucoup d'innovations qu'il convient de suivre, de documenter, d'analyser et de discuter
- Beaucoup d'évolutions qui n'en sont qu'à leurs prémices
- Développer davantage de recherches réflexives sur nos pratiques (données, méthodes, outils, applicatifs...)
- (Re) (Mieux) Positionner la géomatique dans les réflexions autour des données urbaines
  - Mettre à profit nos compétences et nos savoirs pour enrichir les débats

- Importance de l'AED dans la démarche, c'est la méthode préalable avant l'interprétation et la prise de décision
  - Tuckey
- Le continuum sociotechnique comme cadre intégrateur permettant une approche globale de l'objet de recherche et un positionnement à géométrie variable dans la réflexion