




ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES



ined

L'Open Data pour enrichir les données administratives ?

Journée d'étude et d'échanges autour du projet Big_Stat
Arnaud Bringé, Ined, 16 novembre 2021



Plan de la présentation

- Introduction aux données ouvertes
- Comment capter des données ouvertes ?
- Quelques exemples de données disponibles
 - Le site data.gouv.fr
 - Insee, Ined
 - Autres sites officiels
- Quelles utilisations ?
- Problèmes posés

Pourquoi des données ouvertes

- Enrichir les données administratives pour l'analyse :
Par exemple, avec des données contextuelles, pour caractériser des espaces géographiques
- Permettre un cadrage de données
- Opérations de data cleaning
 - Bases de données de références (lieux, prénoms, nomenclatures ...)

Modes d'accès aux données ouvertes

- Fichier de données, téléchargeable
- Visualisation par carte interactive
- Accès via une API dédiée, facilitant un accès sélectif à des données agrégées. Pilotable par logiciel
- Méthodes de scraping : outils dédiés (Rvest, scrapy, extractify ...).
(Cf [séminaire Ined RSA 02.2021](#))

Quels droits ?

- Types de licence :
 - Licence [ETALAB](#)
 - Licence [ODBL](#)
 - Licences [Creative Commons](#)

Problèmes de droit international, et de propriété intellectuelle.

Quelques sites référence

- Le site data.gouv.fr
[Réflexion réutilisation des données ouvertes](#)
- Domaine de la démographie :
 - [Historiques de populations](#) (Insee)
 - [Evènements par commune](#) (Insee)
 - [Bases de données en ligne](#) (Ined)
- Domaine cartographique
 - Le site [Openstreetmap](#)
 - Données de l'[IGN](#)

Modes d'accès (API)

- INSEE : Accès aux
 - Indices et séries chronologiques (BDM)
 - Données locales
 - Métadonnées
 - Répertoire Sirene

- Mise en œuvre assez complexe :
- Création d'un compte
- Abonnement aux API désirées
- Génération de jetons
- Accès possible par R ou SAS

Modes d'accès (API,2)

- Documentations associées :
 - [Catalogue des APIs](#)
 - [Projet UtilitR](#), chapitre 17

Quelques sites référence (santé)

- [DREES](#), portail dédié à l'Open data
- [Cartographie des données de santé publique](#) (Etalab)

Quelques sites référence (Economie)

- Fichier entreprises [Sirene](#) (INSEE)
- Sites [Etalab](#)

Quelques sites référence (Territoires)

- Cartographies municipales : [Paris](#) (licence ODBL), [Lyon](#)
- Logement ([APUR](#), [IAU](#) (licence ODBL))
- Transports ([SNCF](#), [RATP](#) (Etalab, ODBL), [STIF](#), [Transport gouv](#))
- [Equipements communaux](#) (INSEE)

- Données datées, durée de vie limitée
- Données sur un périmètre défini
- Utilisation de données carroyées
- Accès à la documentation associée
- Données internationales

- En cas de données longitudinales :
 - Evolution de la structure communale (Insee)
 - Accès aux données contextuelles anciennes