

Un regard sur l'open data dans les collectivités territoriales

Une étude de l'Observatoire Data Publica en partenariat avec
Opendatasoft

Pourquoi cette étude ?

Plusieurs études sont régulièrement publiées et mises à jour concernant l'open data en France et plus particulièrement l'open data des collectivités territoriales.

La principale et la plus exhaustive est celle de **l'Observatoire de l'open data des territoires** (initiée par Open Data France)

<https://www.observatoire-opensdata.fr/>



Chaque année, le baromètre de **l'Observatoire Data Publica** revient aussi sur ces questions.

<https://observatoire.data-publica.eu/>

D'autres travaux se penchent en détail sur certains enjeux ou plus largement sur l'évolution de l'open data en France. Citons la publication récente de l'ouvrage de **Samuel Goëta** « Les données de la démocratie. Open data, pouvoirs et contre-pouvoirs » aux éditions C&F éditions.

Pourquoi cette étude ?

Aucune étude publique ne s'est penchée jusqu'à présent sur **le détail des données publiées par les collectivités françaises** et sur les mécanismes de publication ou de « moissonnage » des données.

Qui dispose de son propre portail ? Quelles **interactions entre portails** de collectivités et d'administrations centrales ? Quelles interactions entre portails de collectivités ? Quels sont les **jeux de données les plus publiés** ? Pourquoi ? Quels **types de jeux de données** ? Pour quels **usages** ? Quelles informations disposons-nous sur leur **réutilisation** ?

Cette étude s'inscrit aussi dans un contexte : celui de la refonte du projet de l'association Open Data France. Les quelques chiffres qui suivent jettent un regard sur les priorités des collectivités, et donc sur leurs besoins en termes d'accompagnement ou sur des enjeux concrets qu'il conviendrait d'exprimer et de porter collectivement.

Des données issues des portails Opendatasoft



Selon l'Observatoire de l'open data des territoires, environ 1000 acteurs publics territoriaux publient des données ouvertes dont **868 collectivités** (ce qui représente seulement 16% de celles soumises à l'obligation légale de publier leurs données « par défaut »).

Sur ces 868 collectivités, **147 se sont dotées d'un véritable portail open data**, c'est-à-dire une plateforme dédiée, clairement identifiée comme le point d'entrée pour accéder aux données publiques de la collectivité.

Les autres publient sur des portails mutualisés (dont data.gouv.fr) ou pour beaucoup sur des géoportails, avec une limitation du type de données à l'information géographique (à laquelle s'ajoute parfois des données diverses géolocalisées).

Des données issues des portails Opendatasoft



Sur les 147 portails propres de collectivités, 120 utilisent la solution proposée depuis 2015 par la société française Opendatasoft.

Cette position dominante est régulièrement l'objet de commentaires, voire de controverses. Elle doit être nuancée. ODS « équipe » au total seulement 14% des collectivités engagées dans la publication de données. Mais **ODS « propulse » les données de 80% de celles qui ont fait le choix politique et technique de se doter de leur propre portail.**

La liste de ces collectivités est impressionnante. Y figure la quasi-totalité des pionniers de l'open data, avec quelques exceptions notables (la plateforme du Grand Lyon fait office de référence en matière de plateforme open source). Y figurent aussi les collectivités qui publient le plus de données, qui « éditorialisent » le plus leurs portails, et qui les ouvrent plus à des producteurs extérieurs.

Des données issues des portails Opendatasoft



Pour les besoins de la présente étude, il était donc logique et cohérent que l'Observatoire Data Publica se plonge dans les données publiées sur les portails ODS. En effet, ces données, par définition publiques et ouvertes, sont toutes disponibles sur un hub unique <https://data.opendatasoft.com/pages/home/>

Fin 2023, on y trouvait plus de 32 000 jeux de données dont **16 802 jeux de données issus des portails des collectivités françaises**. On notera aussi avec intérêt que **50% des jeux de données des collectivités disponibles sur data.gouv.fr sont à l'origine publiées sur les portails ODS** et « propulsés » de façon automatique.

Ce sont les **16 802 jeux de données** publiés sur des portails propres ODS qui ont servi de matière à la présente étude. Ils constituent à l'évidence un échantillon très large et représentatif des données ouvertes par les collectivités françaises.

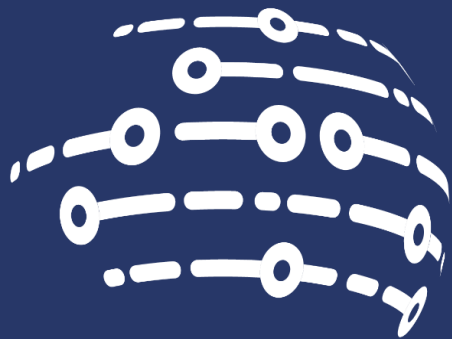
Remerciements et précisions



Les auteurs de la présente étude tiennent à remercier Jean-Marc Lazard, Président d'Opendatasoft, et son équipe pour avoir accepté de jouer le jeu de cette étude et nous avoir parfois aidé à extraire des informations à partir de leur hub.

On notera de façon utile, à l'attention notamment des clients d'Opendatasoft, que cette coopération a porté uniquement sur des éléments publics (données, métadonnées, réutilisations identifiées...) et qu'ODS n'a transmis aucune information supplémentaire à l'Observatoire Data Publica.

Le contenu des portails open data



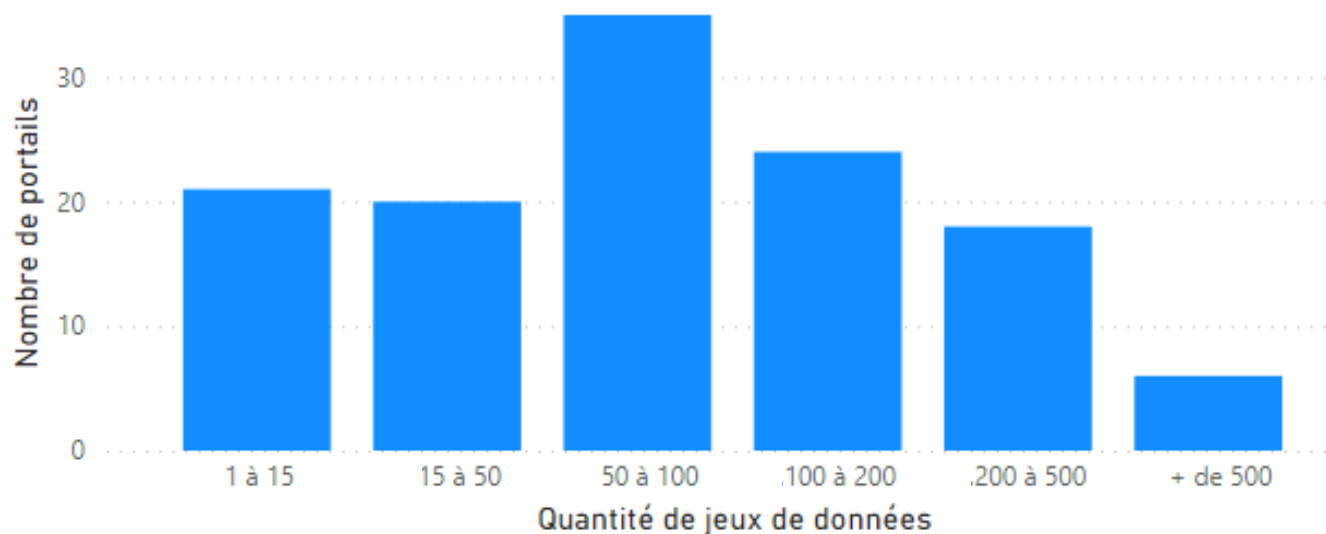
Combien de jeux de données dans les 120 portails ?

138

Nombre de jeux de données
moyen par portail Open Data

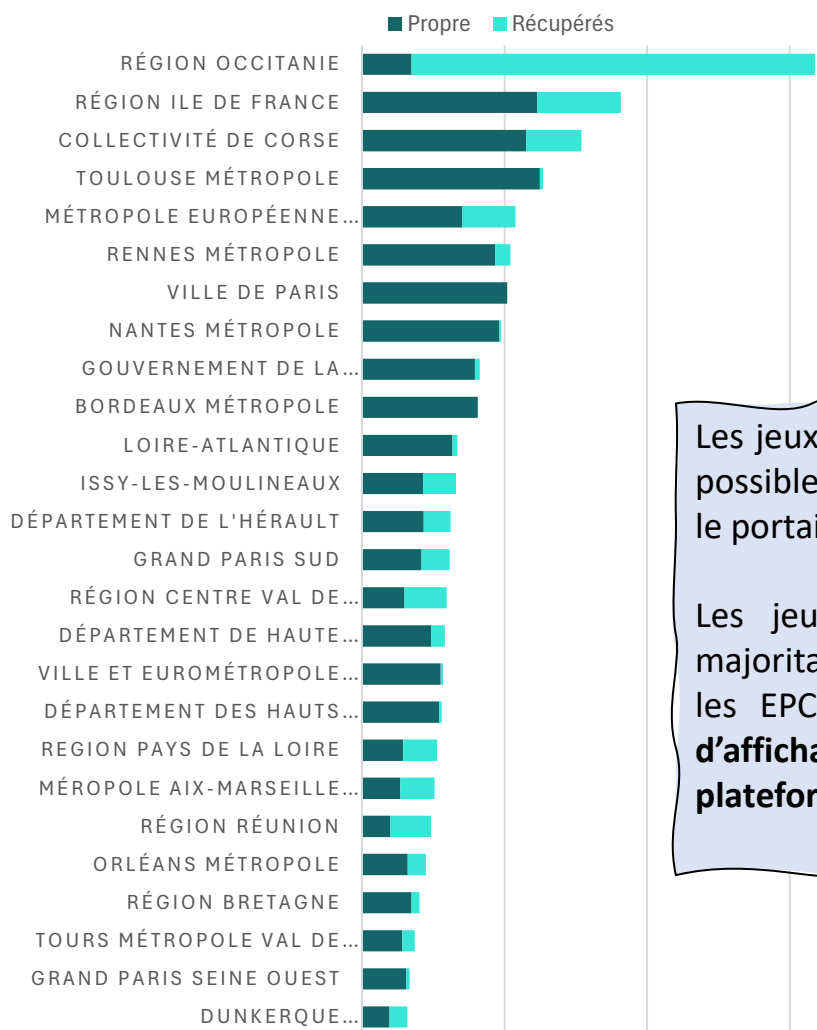
- La moyenne est de 138 jeux de données par portail, mais l'écart est assez important entre les régions et les petites communes.
- A titre d'illustration, **41 portails ont entre 1 et 50 jeux de données** et **49 portails ont entre 50 et 200 jeux de données**.

PORTAILS EN FONCTION DE LA QUANTITÉ DE JEUX DE DONNÉES

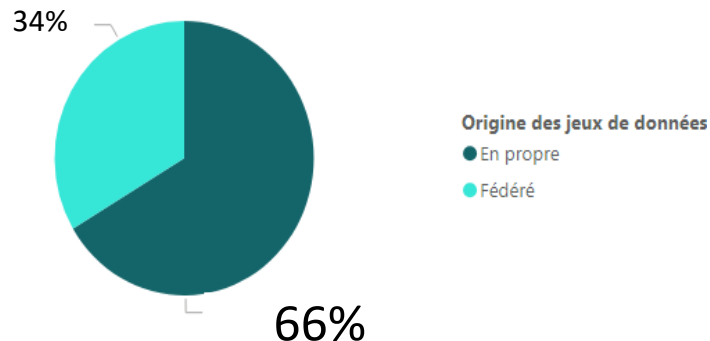


Jeux de données propres ou « moissonnés » ?

JEUX DE DONNÉES PAR PORTAILS ET PAR ORIGINE



PART DES JEUX DE DONNÉES EN PROPRES ET RÉCUPÉRÉS

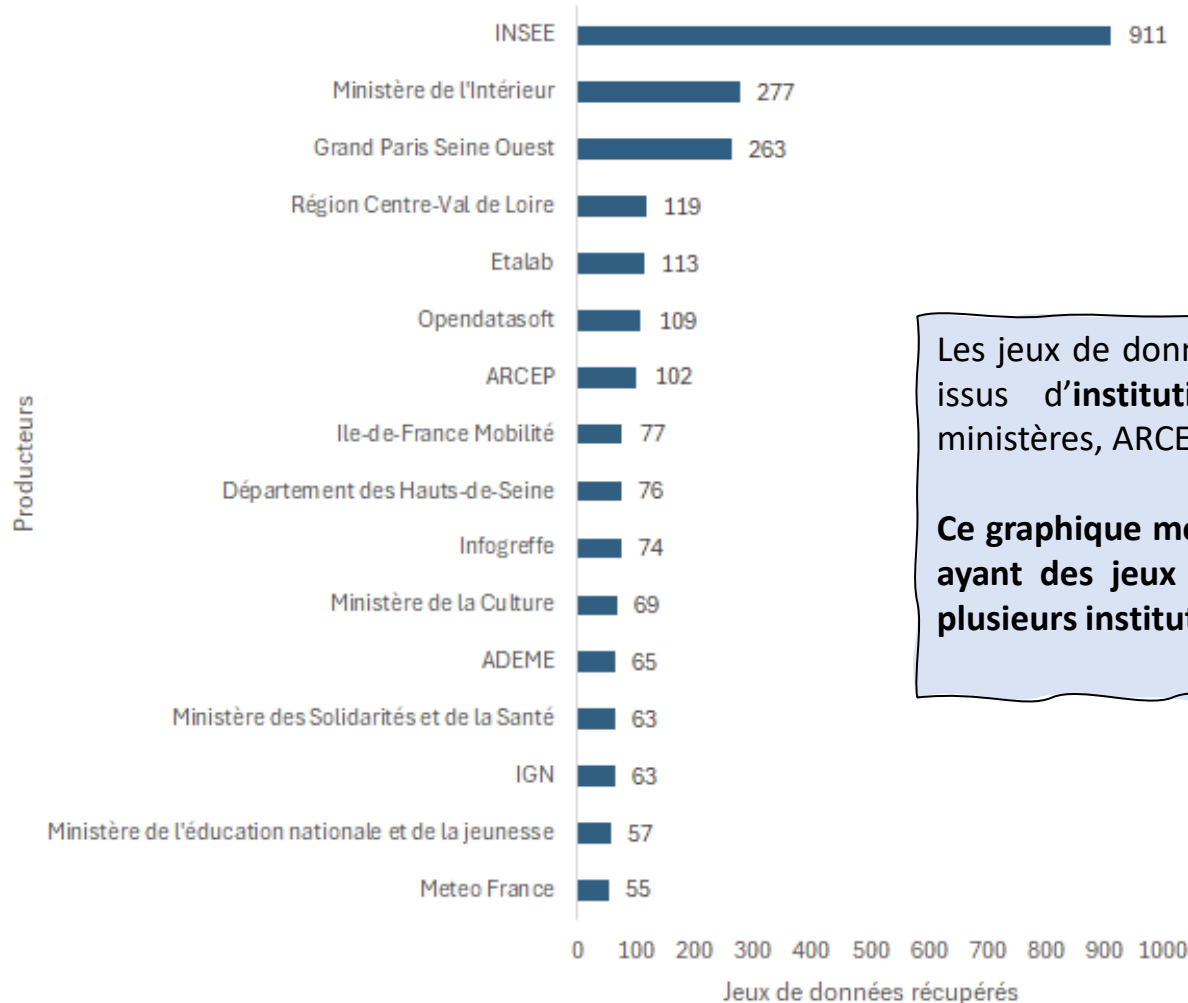


Les jeux de données publiés par les collectivités ont 3 origines possibles : ils sont soit publiés par la collectivité, soit publiés sur le portail par des partenaires, soit « moissonnés ».

Les jeux de données produits par les collectivités sont majoritaires. Mais de plus en plus de jeux sont récupérés par les EPCI et les régions dans une **volonté d'agrégation et d'affichage de l'ensemble des données d'un territoire sur une plateforme unique.**

Qui sont les producteurs les plus « moissonnés » ?

PRODUCTEURS AYANT LE PLUS DE JEUX DE DONNÉES RÉCUPÉRÉS PAR DES PORTAILS

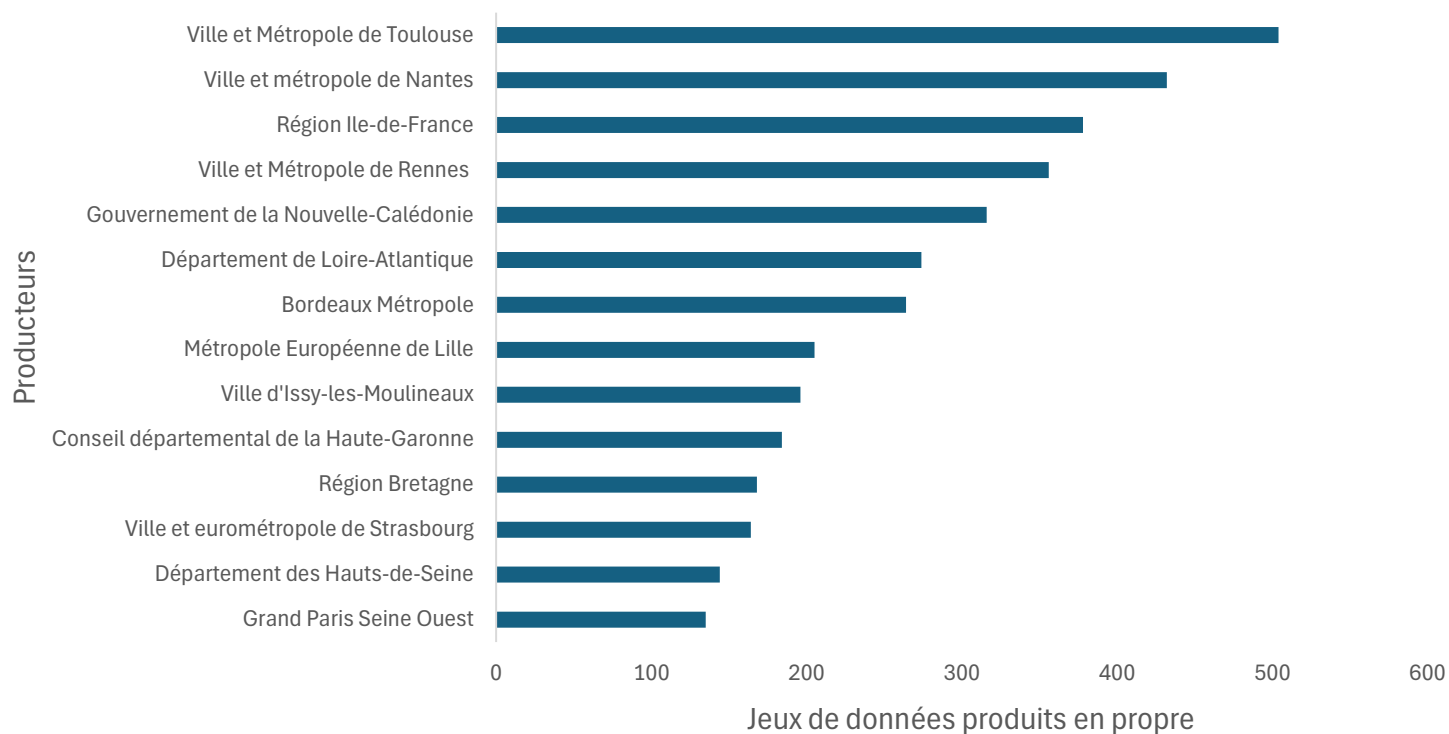


Les jeux de données les plus récupérés sont issus d'**institutions nationales** : INSEE, ministères, ARCEP, Etalab, ADEME...

Ce graphique met en avant les **producteurs** ayant des jeux de données récupérés par plusieurs institutions.

Qui sont les plus gros producteurs de données ?

JEUX DE DONNÉES PRODUITS EN PROPRE PAR PRODUCTEUR

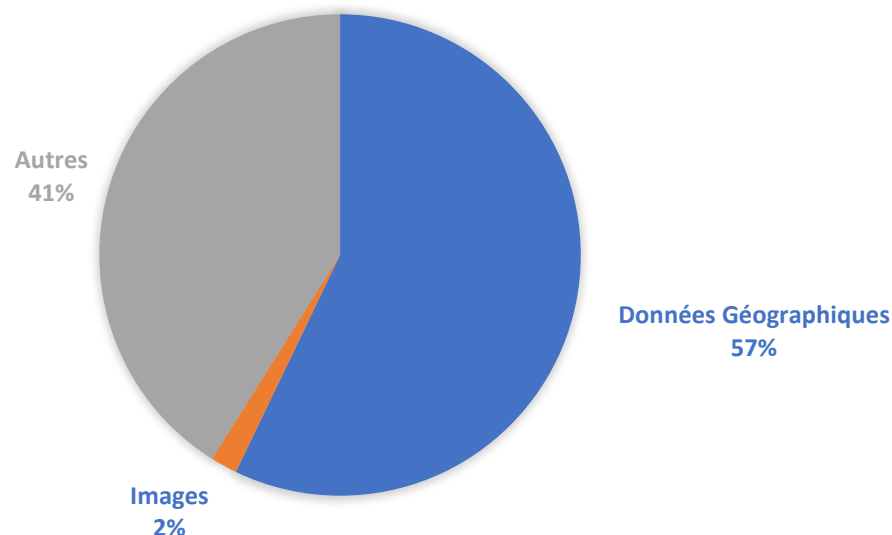


Au sein de notre échantillon de plateformes de collectivités les trois principaux producteurs de données en propre sont :

- Ville et Métropole de Toulouse
- Ville et Métropole de Nantes
- Région Ile-de-France

Les types de jeux de données les plus publiés

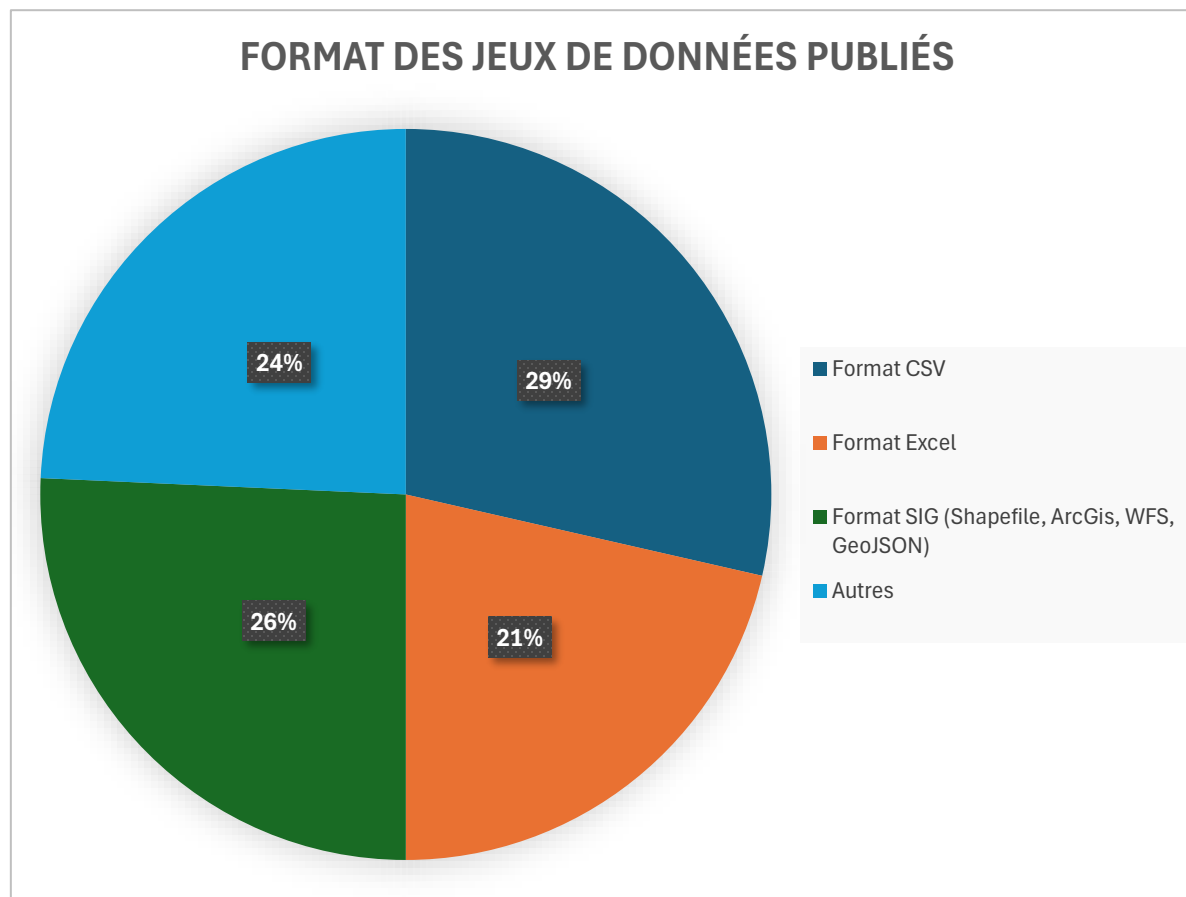
PART DES DIFFÉRENTS TYPES DE JEUX DE DONNÉES



Sur 16 802 jeux de données, les **jeux de données géographiques (SIG)** sont les jeux de données les plus publiés (**57% des JDD**), cela s'explique par l'antériorité de longue date d'un SIG dans de nombreuses collectivités, avec des habitudes et une expertise dans la production de données standardisées et ouvertes.

Dans « autres », on retrouve tout d'abord des données chiffrées (budgets, résultats des élections...), des données textuelles (délibérations par exemple) ou parfois des jeux de données mixtes (schémas et plans d'urbanisme notamment). A noter : les images commencent à faire partie des données publiées.

Les formats des données publiées



Le CSV est la star incontestée de l'open data... avec le format tableur propriétaire Excel !

50% des jeux de données sont donc des tableaux. A noter : on retrouve en tableau des données de familles très diverses, y compris de nombreuses coordonnées géographiques.

Le « socle commun des données locales »

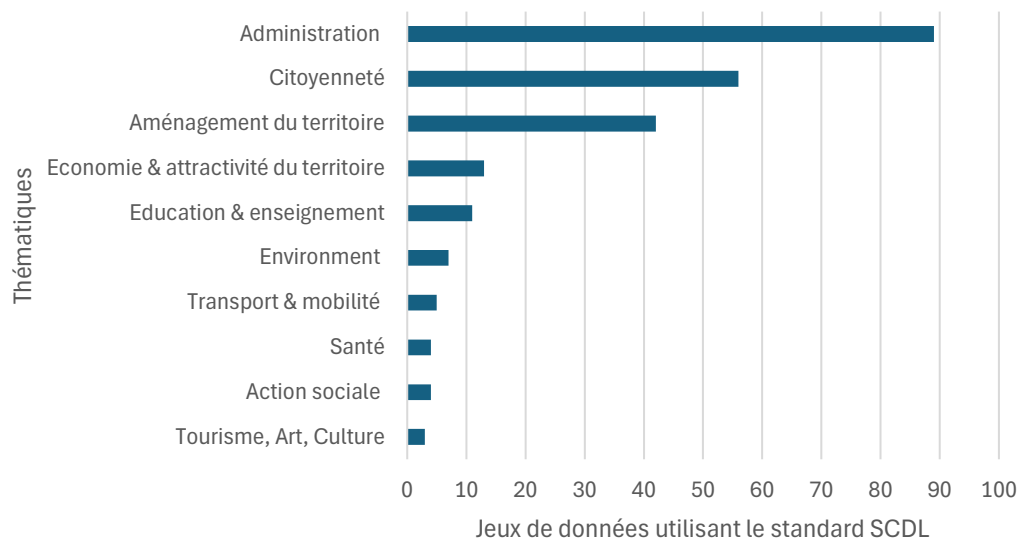
La réutilisation des données ouvertes est freinée par le manque d'homogénéité des données et des difficultés d'interopérabilité. En réponse à ces enjeux, OpenDataFrance a engagé en **2017** une réflexion sur la standardisation des données jugées les plus essentielles. La démarche OpenDataLocale a débouché sur la création d'un « **Socle Commun des Données Locales (SCDL)** ».

Ce standard identifie les jeux de données prioritaires qui doivent suivre un formalisme commun :

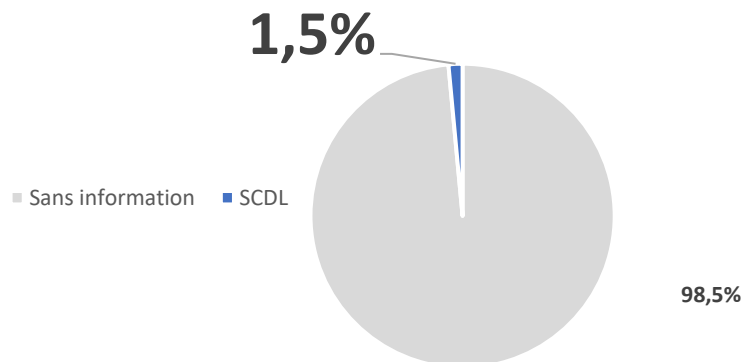
- Aménagements cyclables
- Base Adresse Locale
- Budget des collectivités et établissements publics locaux
- Catalogue de données simplifié
- Composition des plats des menus et menus de la restauration collective
- DAE (localisation géographique et caractéristiques techniques des Défibrillateurs Automatisés Externes (DAE) accessibles au public)
- Délibérations
- Equipements
- IRVE (infrastructure de recherches pour véhicules électriques)
- Prénoms des nouveaux-nés
- Subventions

Le « socle commun des données locales »

JEUX DE DONNÉES UTILISANT LE STANDARD SCDL PAR THÉMATIQUE



PART DES JEUX DE DONNÉES UTILISANT LE STANDARD SCDL



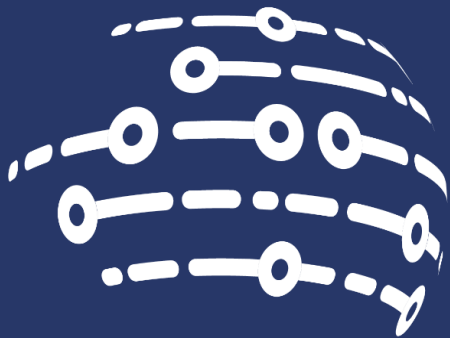
Force est de constater que ce standard n'a pas encore complètement rencontré son public... ou qu'il convient d'étendre la démarche à d'autres familles de données.

Seulement 1,5% des jeux de données publiés identifient le standard SCDL.

Une nuance cependant : les chiffres exposent l'ensemble des jeux de données identifiés par les collectivités propriétaires comme respectant le standard SCDL.

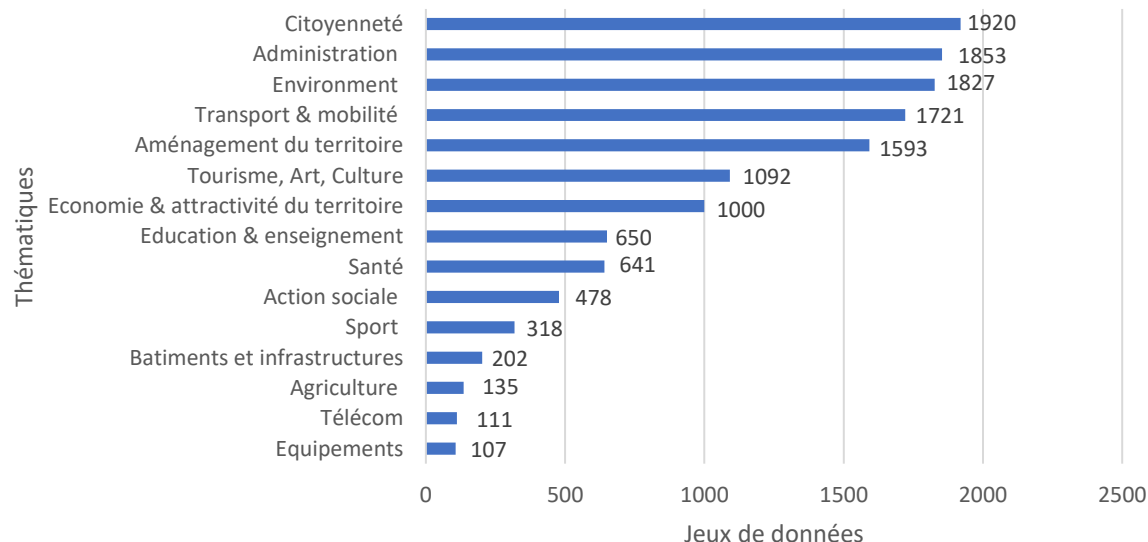
La capacité des acteurs à identifier les standards et schémas utilisés est un enjeu collectif stratégique.

Les jeux de données



Les thématiques phares

JEUX DE DONNÉES PAR THÈME

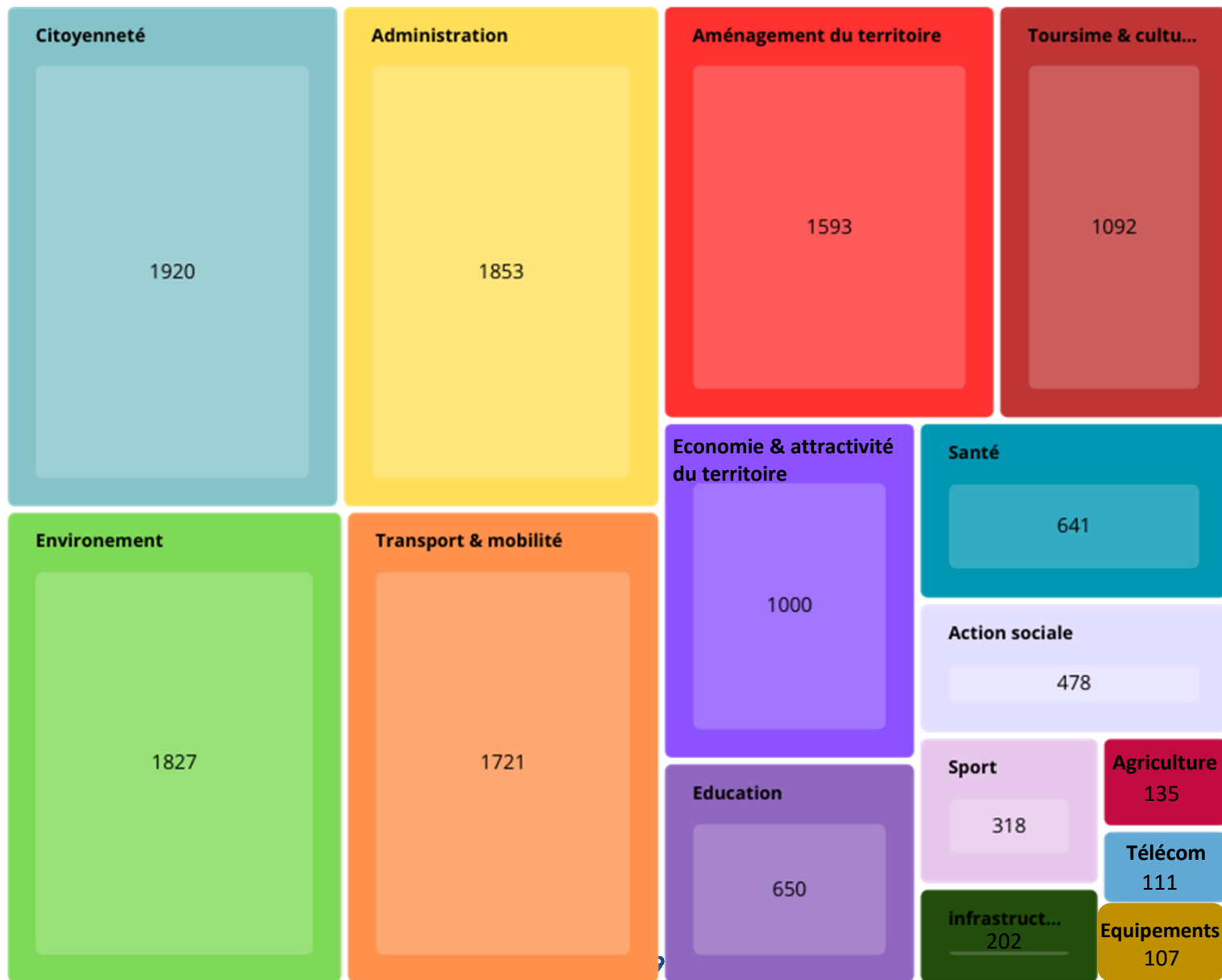


Le TOP 3

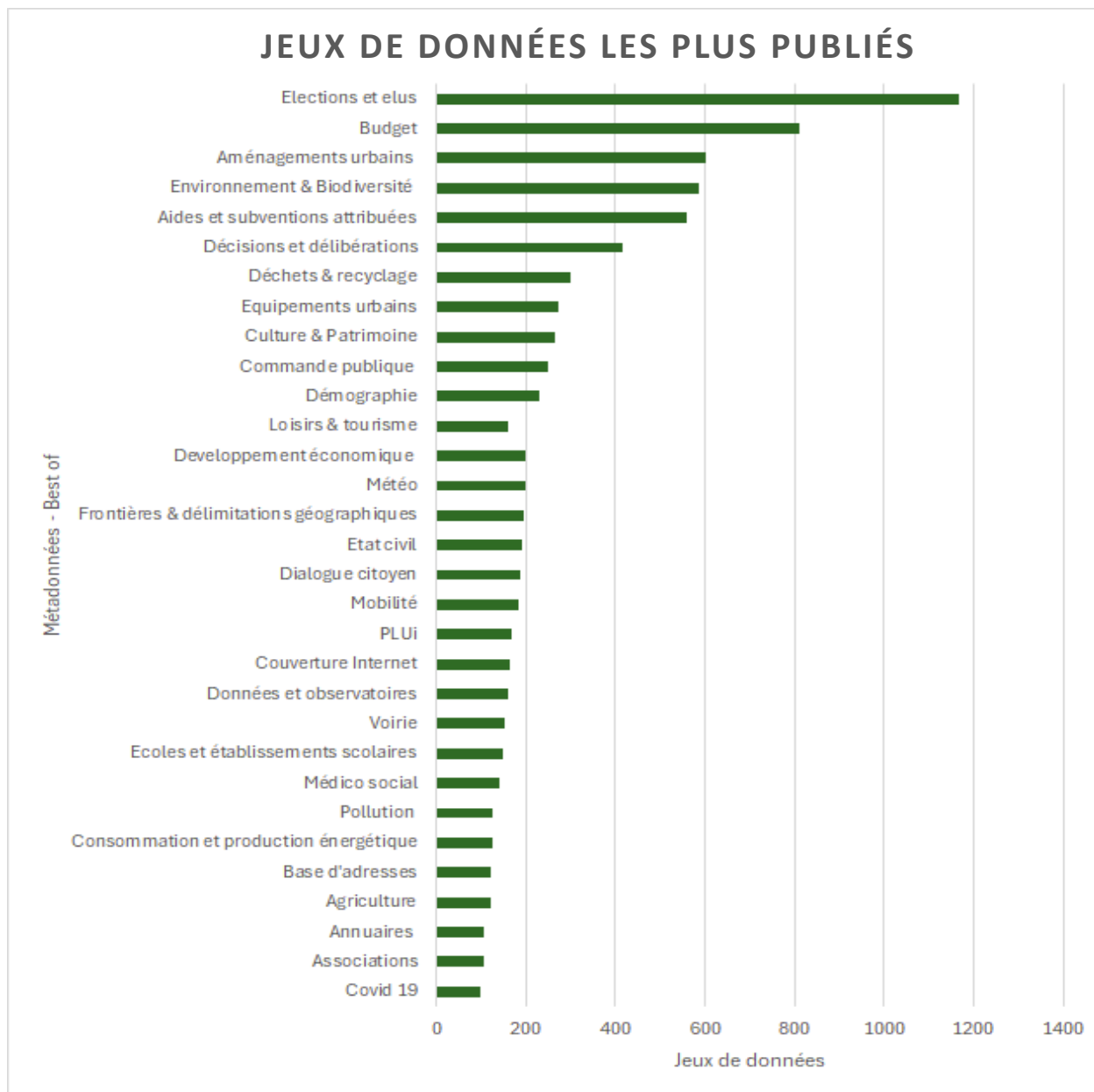
- **Citoyenneté** : informations relatives aux services publics et aux dispositifs électoraux (bureaux de vote, circonscriptions, participation citoyenne, institutions, vie citoyenne...).
- **Administration** : documents administratifs produits par la collectivité (budgets, délibérations, commande publique, services d'état civil, vie de l'institution...).
- **Environnement** : données relatives à l'eau (assainissement, cours d'eau...), la biodiversité (réserves naturelles, natura 2000...), les conditions climatiques (qualité de l'air, précipitations...) et la transition écologique (PCAET, énergie, déchets...)

Les thématiques n'étant pas toujours identifiées, un travail conséquent de mise en correspondance des thèmes a été nécessaire.

Les thématiques phares



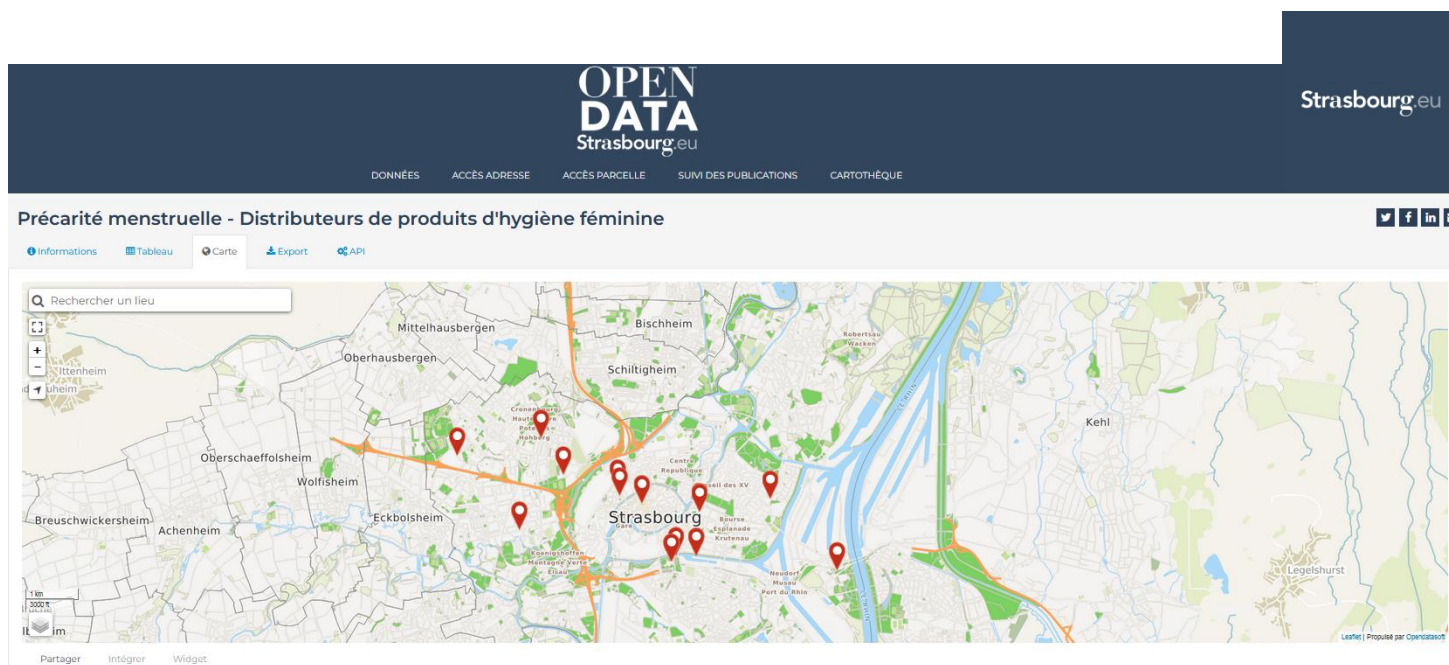
Les jeux de données les plus publiés



Quelques jeux de données atypiques

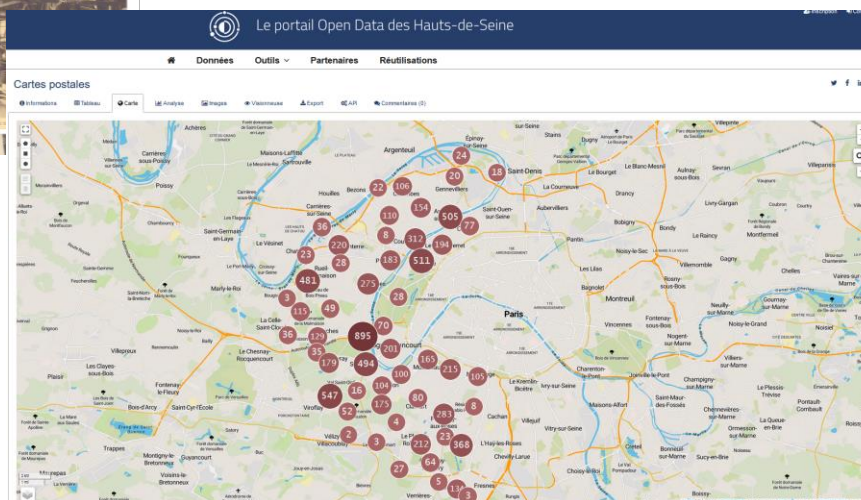
Au fil de notre analyse, nous avons trouvé quelques jeux de données atypiques

- **Jeu de données identifiant les distributeurs de produits d'hygiène féminine**



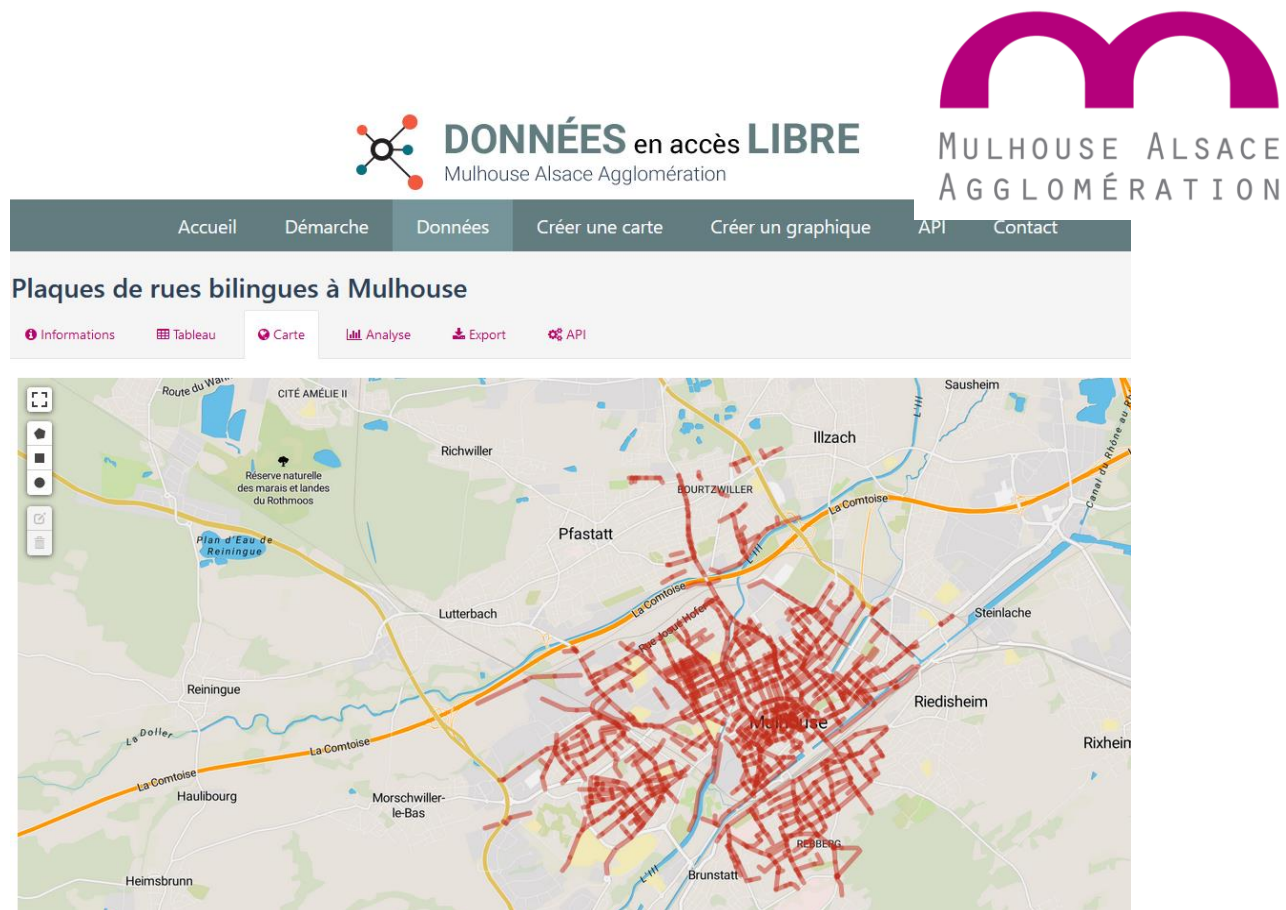
Quelques jeux de données atypiques

- Jeux de données sur les cartes postales



Quelques jeux de données atypiques

- Jeu de données des plaques de rue bilingue



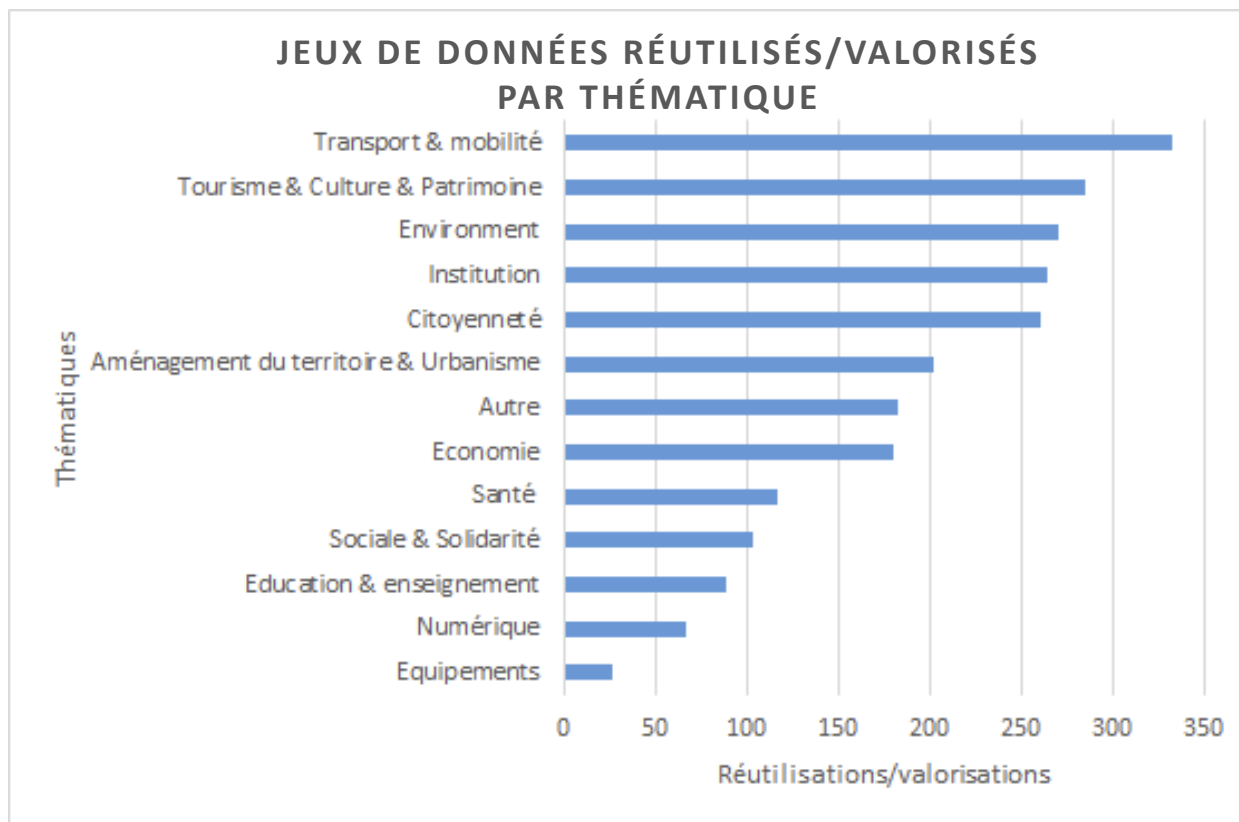
La réutilisation et la valorisation des données



Les réutilisations par les collectivités

Les premiers réutilisateurs des données sont... les collectivités elles-mêmes !

En effet, à partir des données qu'elles ouvrent et mettent en ligne, elles produisent de nombreuses datavisualisations, cartes interactives, graphiques et outils d'analyse. Certains y verront un dévoiement de l'open data qui devient « communication », d'autres au contraire salueront le travail d'accessibilité accompli. Mais dans tous les cas de figure, le jeu de données brutes est accessible.



Les réutilisations par les collectivités

TOP 3

Le **top 3** des jeux de données les plus réutilisés directement par les collectivités (en dataviz ou en carte) regroupe les **thématiques de la mobilité, du tourisme et de la citoyenneté**.

Les **principales réutilisations** se trouvent dans le domaine de la **mobilité** notamment à travers la production de **cartes** (points de comptage des véhicules, circuits de randonnées en vélo, tracés des pistes cyclables) ou d'**applications** (détection de place de stationnement libres, localisation de parkings, transports en commun...).

Dans le domaine du **tourisme et de l'attractivité territoriale**, les réutilisations sont également nombreuses. On peut notamment citer les **cartes relatant les circuits de randonnées, l'emplacement des éléments de patrimoines remarquables**.

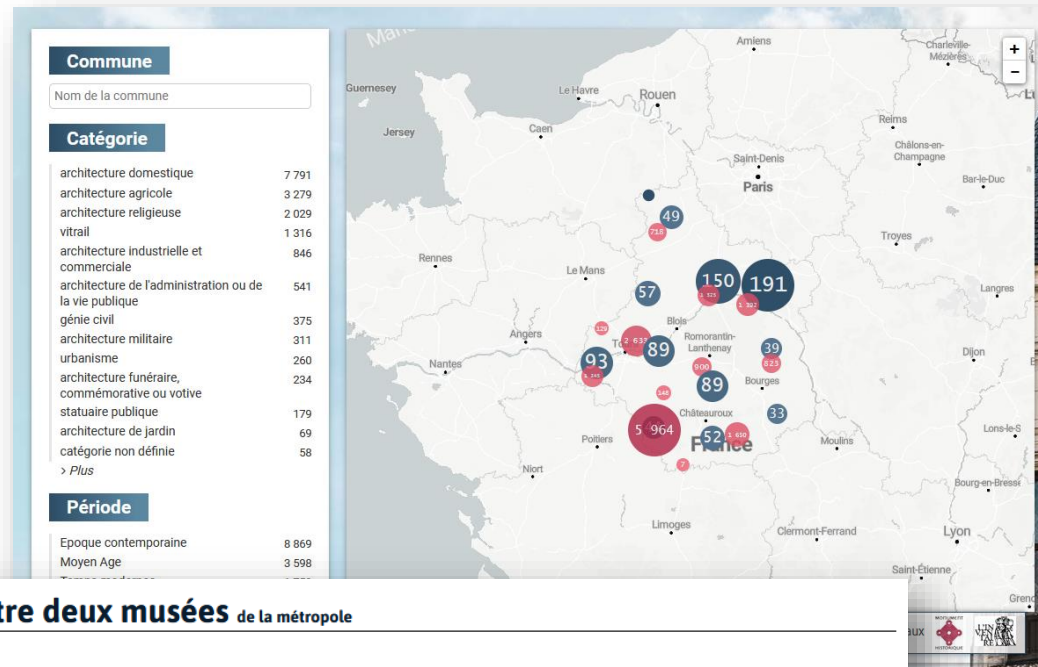
A gauche :
Portail Open data de la ville
de Paris

A droite :
Portail Open data de
Bordeaux métropole

Les réutilisations par les collectivités – Tourisme et patrimoine

*A droite :
Portail Open data de la
Région Centre-Val-de-Loire*

*En bas :
Portail Open data de la
Métropole Aix-Marseille-
Provence*



Comparer la fréquentation entre deux musées de la métropole

Sélectionner un
premier musée

Cabinet des Monnaies et Médailles de Marseille



3 746

Nombre total d'entrées

Sélectionner un
deuxième musée

Musée d'Archéologie Méditerranéenne



1 661 549

Nombre total d'entrées

Les réutilisations par les collectivités – citoyenneté

A gauche :
Portail Open data de la ville
de Paris

985 enregistrements

Aucun filtre actif

Filtres

Rechercher...

mandature

1989-1995	177
2008-2014	175
2001-2008	172
1983-1989	171
1995-2001	170
1977-1983	120

conseiller

Conseiller	693
Conseillère	292

secteur

15ème arrondissement	102
18ème arrondissement	86
13ème arrondissement	77
16ème arrondissement	77

Les membres du Conseil de Paris depuis 1977

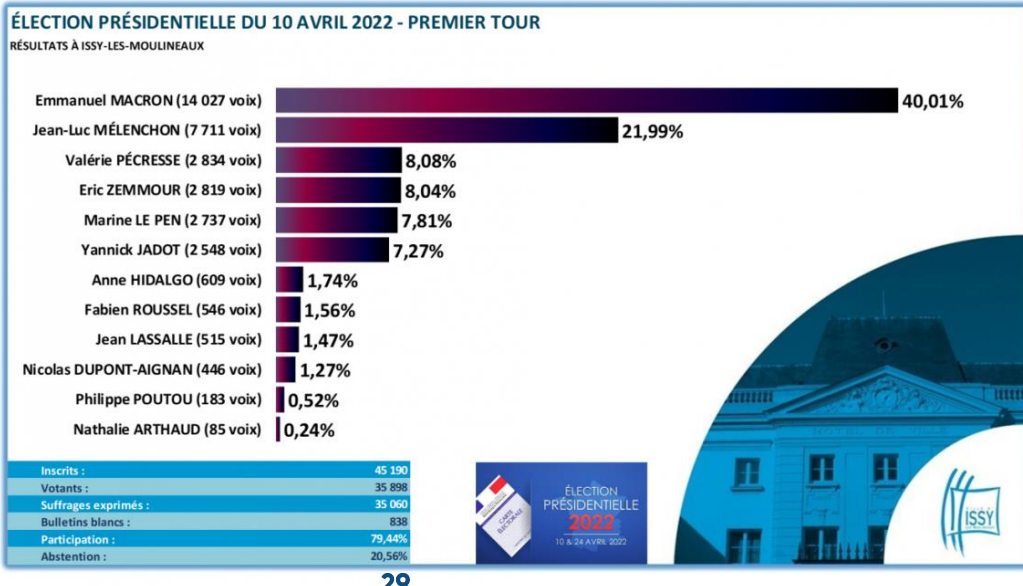
De nouveaux enregistrements sont actuellement en cours d'ajout sur ce jeu de données. Pendant la durée du traitement, les visualisations peuvent afficher des données incomplètes. N'hésitez pas à [rafraîchir votre page](#) régulièrement !

Informations Tableau Analyse Images Export API



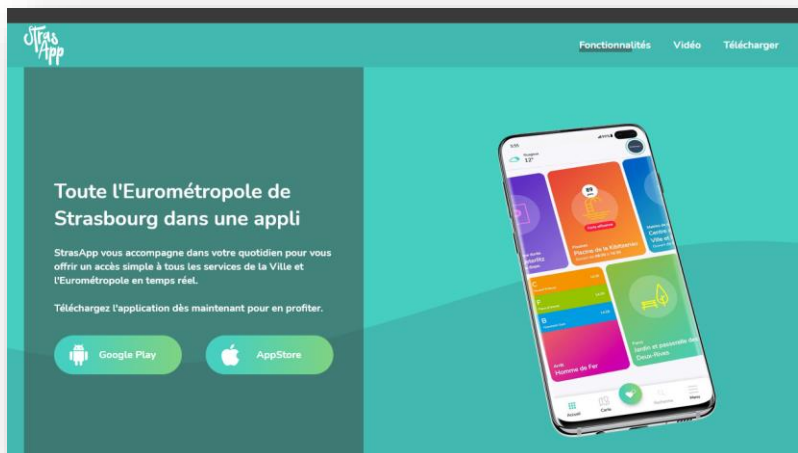
Partager

A droite :
Portail Open data de la ville
d'Issy-les-Moulineaux



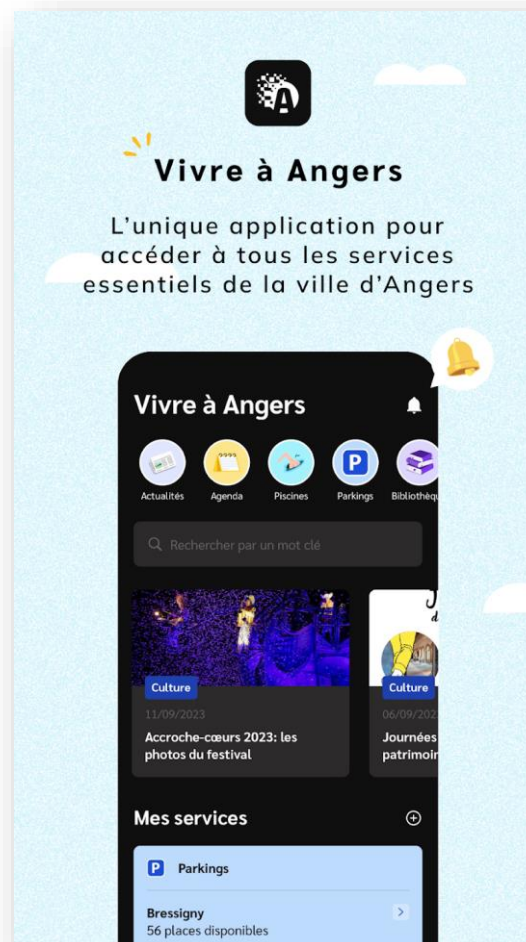
Les réutilisations par les collectivités – citoyenneté

Stras App, l'application de la Ville et l'Eurométropole de Strasbourg et Vivre à ANGERS, l'application de la ville d'Angers, utilisent directement des données en open data pour remettre les informations à disposition du grand public.



<https://strasapp.eu/>

*Capture d'écran de
l'application « Vivre à
Angers »*



Les téléchargements : la mobilité largement en tête !

Bien sûr, les réutilisations sont aussi le fait de partenaires ou d'acteurs indépendants qui téléchargent les jeux de données. Mais l'un des principes fondateurs de l'open data (le « libre accès aux données », donc sans identification) rend impossible le « traçage » de nombreux usages, sauf si ceux-ci sont proposés par les réutilisateurs eux-mêmes, via un formulaire par exemple les [réutilisations de Bordeaux](#), les [datavisualisations hébergées sur le site internet de la région Ile de France](#).

A défaut d'inscription, il est intéressant de comptabiliser les téléchargements. Les **jeux de données des collectivités territoriales qui sont les plus téléchargés** sont les **jeux de données de la mobilité**. Ces données alimentent de nombreux usages au premier rang desquelles les applications mobiles. Les chiffres sont parfois impressionnants, ce qui s'explique par des mises à jour fréquentes, quand ce n'est en « temps réel ».

Quelques exemples recensés (valeurs décembre 2023, chiffres depuis la publication du jeu de données) :

- V'Lille - Disponibilité en temps réel, jeux de données de la Métropole Européenne de Lille – **7,9 millions de téléchargements**
- Etat du trafic en temps réel, jeux de données de Rennes Métropole – **4,3 millions de téléchargements** , et position des bus en circulation sur le réseau STAR en temps réel – **2,7 millions de téléchargements**
- Disponibilité dans les Parking à Angers, jeux de données d'Angers Loire Métropole - – **3,1 millions de téléchargements**

Les réutilisations par des partenaires

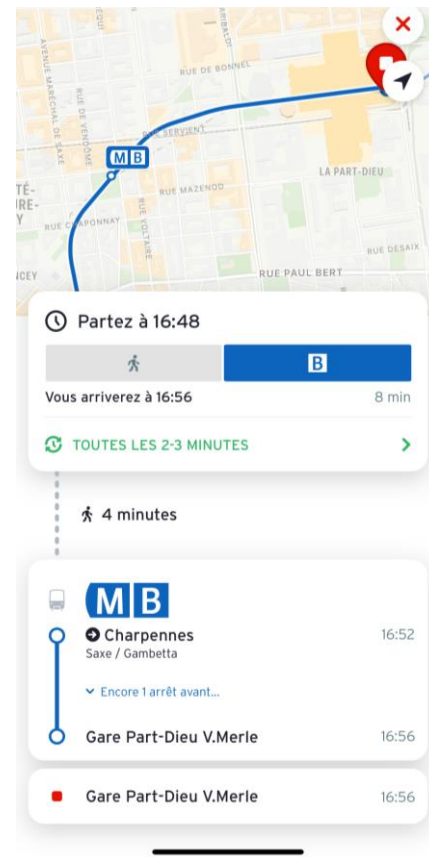
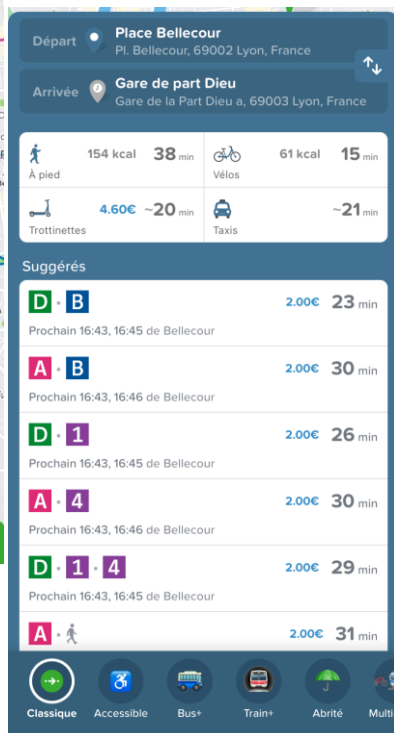
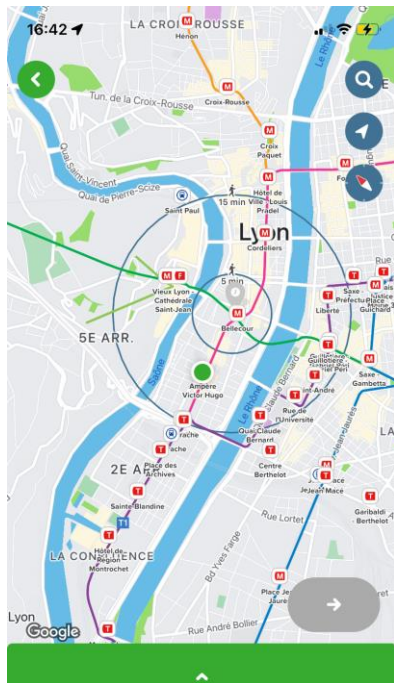
Park'in, l'application utilisant la fréquentation des parkings en temps réel à Bordeaux afin d'en indiquer la disponibilité.

iPhone Screenshots

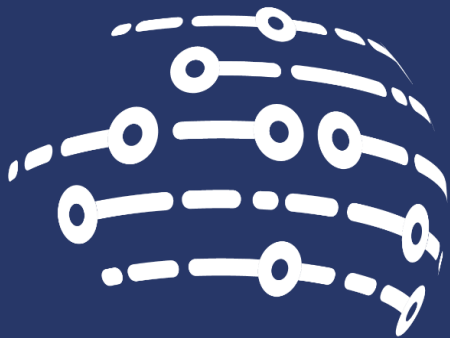


Les réutilisations par des partenaires

Citymapper et Transit, les applications utilisant les data de la métropole de Lyon afin de faciliter les déplacements en transport en commun



Les liens entre portails



Les relations entre portails



Opendatasoft a développé pour ses clients une série d'interfaces automatisées permettant à la fois de récupérer (« moissonner ») des jeux de données qui concernent le territoire et de « propulser » les données de la collectivité vers d'autres portails.

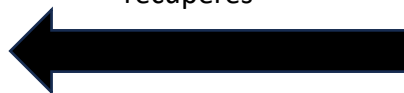
L'Observatoire Data Publica a pu analyser ces flux.

Récupérer des données nationales



90 portails de collectivités

Jeux de données
récupérés



22 portails d'administrations nationales

22 portails d'administrations nationales voient leurs jeux de données récupérés (« moissonnés » ou « fédérés ») par 90 portails de collectivités territoriales.

Le Top 3 des jeux de données automatiquement récupérés par les collectivités territoriales sont :

- #1 : Les **données essentielles de la commande publique** publiées sur le portail du Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique <https://data.economie.gouv.fr> (57 liens)
- #2 : L'annuaire de l'éducation du ministère de l'Éducation nationale, de la Jeunesse et des Sports <https://data.education.gouv.fr/explore/dataset/fr-en-annuaire-education/information/> (49 liens)
- #3 : La liste des immeubles protégés au titre de monuments historiques du Ministère de la Culture <https://data.culture.gouv.fr/explore/dataset/liste-des-immeubles-proteges-au-titre-des-monuments-historiques/information> (20 liens)

Partager des données entre collectivités

Les liens entre les portails de collectivités se font de plusieurs manières :

- La distribution de jeux de données



La distribution de jeux de données est le fait pour un portail d'une collectivité "parente" de pousser automatiquement les données vers une collectivité de la strate inférieure. Ainsi, des régions distribuent des données vers les EPCI ou les départements et les départements/EPCI en distribuent vers leurs communes.

Partager des données entre collectivités

- Le moissonage automatique de jeux de données



La récupération de jeux de données par “fédération” ou moissonage se fait de deux manières différentes :

- En tant qu'agrégatrices de données, les collectivités telles que les EPCI et les régions vont récupérer les données des portails sur leur territoire.
- Les collectivités de strate inférieure vont récupérer les données de la strate supérieure pour alimenter leur portail

**A partir des 120 portails étudiés,
peut-on définir un portail idéal ?**

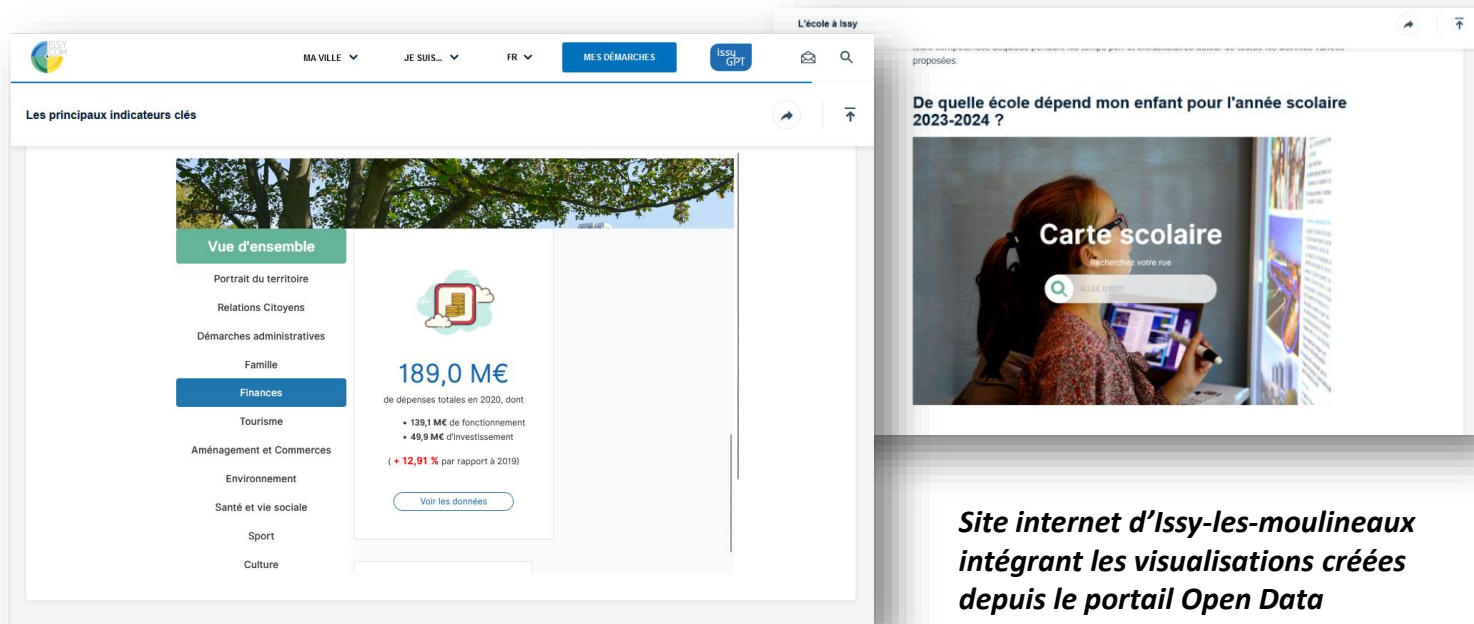


Le portail Open data idéal ?

- Une page d'accueil explique la démarche et met en avant des informations utiles :
 - ✓ Les **dernières données** ajoutées
 - ✓ Un **zoom** sur les **données intéressantes du moment**
 - Ex : La mise en avant des points de collecte de sapins lors des fêtes de Noël
- Un **portail ergonomique** et simple à utiliser
 - ✓ Des **onglets compréhensibles et pertinents**.
 - Des choix de thématiques prioritaires (“éditorialisation” du portail)
 - La “donnée du mois” qui colle à l’actu du territoire
 - Des données clefs du territoire
 - ...
 - ✓ Des **réutilisations mises en avant**
 - ✓ Un **catalogue** de jeux de données ergonomique
 - Des thématiques bien définies
 - Des mots clés bien identifiés
 - Des jeux de données bien qualifiés
 - Des jeux de données mis à jour régulièrement

Le portail Open data idéal ?

- Un **portail lié au site internet de la collectivité**. Il est par exemple intéressant de lier la communication institutionnelle aux jeux de données qui viennent étayer la décision ou la présentation des politiques publiques.



Site internet d'Issy-les-moulineaux intégrant les visualisations créées depuis le portail Open Data

- Un **portail connu et utilisé en interne par les agents** de la collectivité
- La **mise en avant de visualisations** sur des **espaces externes** au portail via un lien rapide

« Coup de cœur »



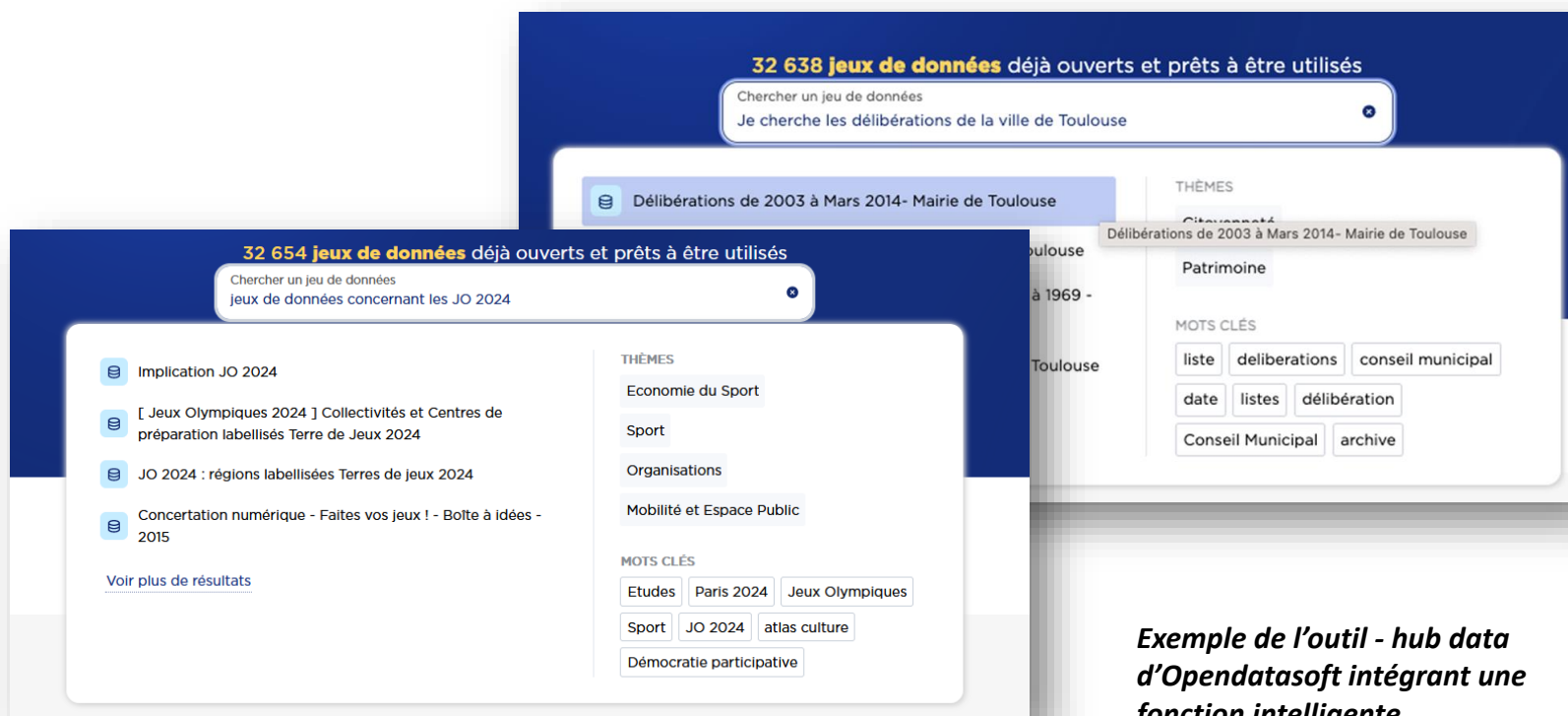
Un portail récent avec un design réussi : Dijon métropole
<https://data.metropole-dijon.fr>



Des moteurs de recherche propulsés par l'IA ?

Pour faciliter la recherche de jeux de données et accroître la « découvrabilité » des données, certains misent sur l'intelligence artificielle. Un moyen de pallier au manque de précision des métadonnées ? Un moyen de contourner l'absence de standardisation ?

A suivre !



Exemple de l'outil - hub data d'Opendatasoft intégrant une fonction intelligente.
<https://data.opendatasoft.com>

Contact

Pour tout renseignement :

observatoire@data-publica.eu

06 74 52 69 23

<https://observatoire.data-publica.eu/>

