

The future of open data portals

THE CHALLENGES OF BUILDING AN USER-CENTRED OPEN DATA PORTAL

Mathilde Hoang

Antonin Garrone

Mario Restuccia

Tuesday 15 December 2020

Who we are



Mathilde Hoang

Open Data Coordinator



Antonin Garrone

Data Editor



Mario Restuccia

Open Data Officer

MISSION

What is Etalab

We coordinate the design and implementation of the State's data strategy.

What is Etalab

- Data **opening** and data circulation
- Data **exploitation**: data science, artificial intelligence
- Innovation and openness of **public action**



PROGRAMME

Our program today



Producers

**Central and
local
administrations**

Any other
publisher



Reusers

Administrations,
companies,
datajournalists...



Curious visitors

People looking
for any kind of
information

**Responding to
producers' needs**

Who are the producers?

- Centralised public administration
- 18 regions, 93 departments, 34 967 communes
- Legal obligation of **open data by default**
- Many more **public and private actors** who create data

 **lots of data**

PRODUCERS

MISSION

What are we trying to do?

One team, multiple skills



managing **data.gouv.fr**



legal support



foster an open data culture



tech support



empowering producers

What challenges for us?

- ⚠ Manage lots of data
- 👤 Allow everyone to publish on data.gouv.fr
- 👩 Make producers autonomous
- 📘 Ensure a high quality of data

**How we are
doing it?**

**First need:
publishing data
our infrastructure**

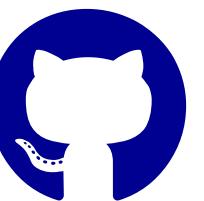




U DATA

Customizable and skinnable social
platform dedicated to (open)data

The platform: udata

- Open source project started by Etalab in 2011
- In 2016 moved to a separate Github organisation  opendatateam
- Currently used by    



data.gouv.fr

Data Reuses Organizations Dashboard Documentation News

Open platform for French public data

Mario Restuccia



Search

Agriculture et Alimentation

Culture, Communications

Comptes, Économie et Emploi

Éducation, Recherche, Formation

International, Europe

Environnement, Énergie, Logement

Santé et Social

Société, Droit, Institutions

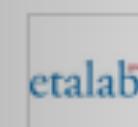
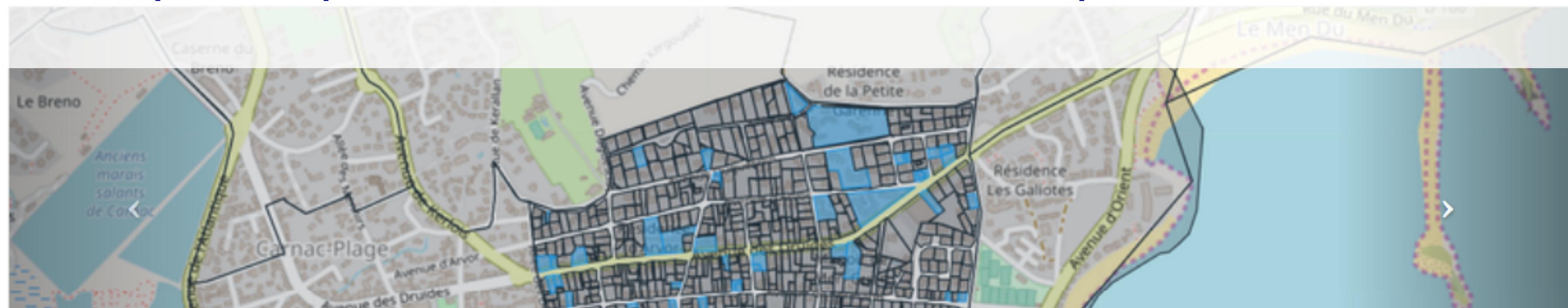
Territoires, Transports, Tourisme

Share, improve and reuse public data

CONTRIBUTE!

data.gouv.fr

Open platform for French public data



ETALAB

EXPLORATEUR DE DONNÉES DE VALEUR FONCIÈRE (DVF)

PRODUCERS

INFRASTRUCTURE

data.gouv.fr

35 000+

Datasets

2 400+

Reuses

2 500+

Publishers

PRODUCERS

INFRASTRUCTURE

How to publish data



direct upload



harvesting

Manually upload files

Resources are
hosted on our
servers

✓ very easy
! have to do
everything by hand

Set up a harvester
for another portal

Supports DCAT/
CKAN/ODS

✓ set it up once
and forget about it



API

Do (almost)
everything via API

✓ powerful
! need tech skills

Going further

support and quality



PRODUCERS

DATA LITERACY

Guides & docs

 guides.etalab.gouv.fr

Aimed at **chief data officers, lawyers, project managers, or just users**

Open source, of course
 [etalab/guides.etalab.gouv.fr](https://github.com/etalab/guides.etalab.gouv.fr)

 doc.data.gouv.fr

Documentation on **how to publish data** on the portal

Also open source
 [etalab/doc.data.gouv.fr](https://github.com/etalab/doc.data.gouv.fr)

Metadata monitoring

When creating a dataset on data.gouv.fr, producers are **encouraged** to provide metadata

Quality



The aim of that box is to help you improve the quality of the (meta)data associated to your dataset.

It gives you an overview of what will be useful for contributors to find and reuse your data.

! Description

! Tags

✓ Open formats

? Discussions

? Up-to-date

✓ Availability

Data schemas

- A schema is a **blueprint** of how the data should be structured
- They allow an increase in the **quality** of the data, especially when several data producers create the same kind of data

PRODUCERS

DATA QUALITY

Utilities for data schemas

 schema.data.gouv.fr

**Reference public
schemas**
justified by law or by
usage

14 schemas today
+ 10 work in progress

 validata.etalab.studio

**Check if data
conforms to a
schema**

A shared project
among different
actors
 [git.opendatafrance.net/
validata](https://git.opendatafrance.net/validata)

 csv-gg.etalab.studio

**Generate a csv
that conforms to a
schema**

 [etalab/csv-gg](#)

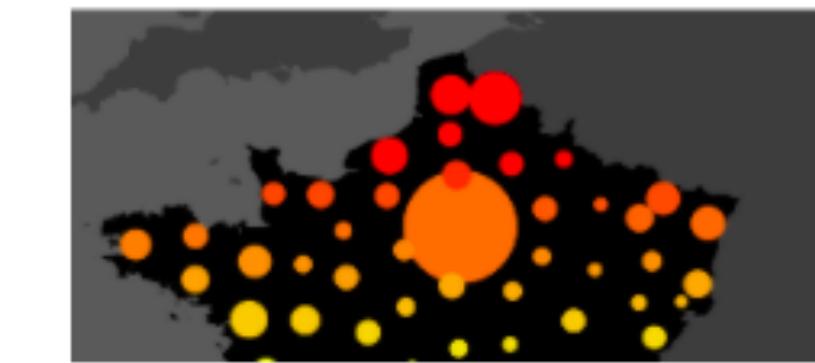
**Responding to
reusers' needs**

REUSERS

Who are the reusers?

Anyone can publish

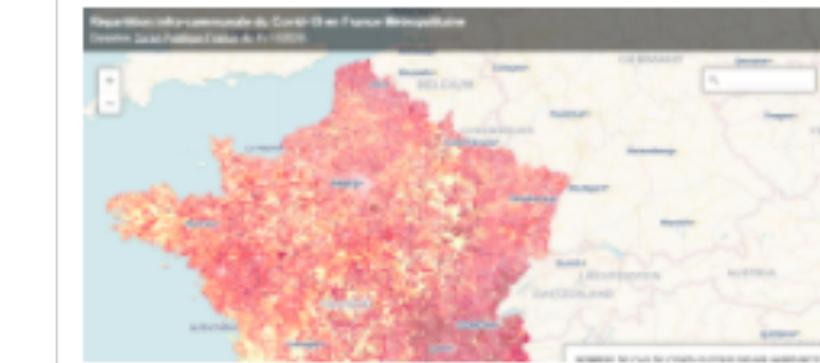
Administrations, private companies, datajournalists...



COVID-19_tests_virologiques

Cartographie interactive automatique des résultats des tests communiqués par les laboratoires , par EPCI et par

[Application](#) 1 ★ 0 [2020](#)



Répartition infra-communale du COVID-19 en France Métropolitaine

Cartographie à l'échelle IRIS du taux d'incidence de l'épidémie de COVID-19 en France Métropolitaine (mise à

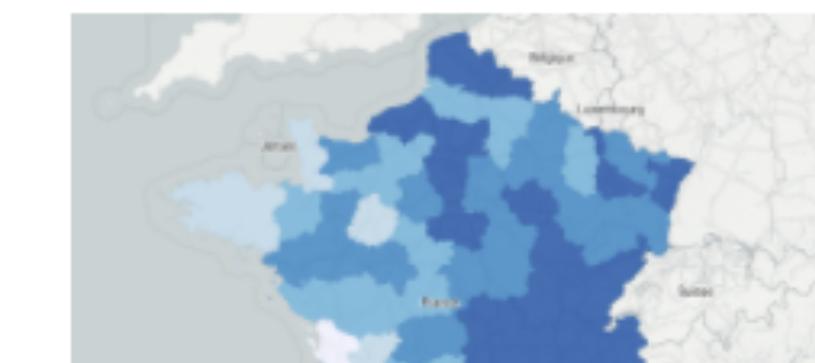
[Visualisation](#) 1 ★ 0 [2020](#)



Taux d'incidence, dépistage et positivité COVID par Commune sur une semaine glissante en France Métropolitaine

Visualisation des taux d'incidence, dépistage et positivité COVID par

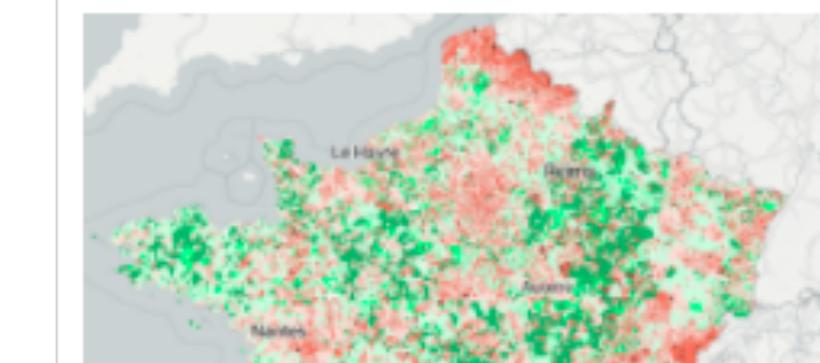
[Visualisation](#) 1 ★ 0 [2020](#)



Observatoire GMS COVID-19

Global Map lance l'observatoire GMS COVID-19 qui permet de suivre l'évolution de la pandémie d'une

[Article de blog](#) 1 ★ 0 [2020](#)



Cartographie COVID-19 au niveau infra-communal

Découvrez en exclusivité la première cartographie liée au COVID-19 au niveau infra-communal. Ces

[Article de blog](#) 1 ★ 0 [2020](#)



Taux d'incidence, de dépistage et de positivité COVID par EPCI sur une semaine glissante en France Métropolitaine

Cartographie interactive en fonction du temps du suivi du taux d'incidence, du

[Visualisation](#) 1 ★ 0 [2020](#)

REUSES

What are reuses?

2 400+

Reuses

Visualisation

Application

Post

api

API

news_article

News
articles

paper

idea

hardware

Legal framework

List of licences set by decree including our own **Open Licence 2.0**

Allowing reproduction, redistribution, adaptation and commercial exploitation of the data

Few requirements :

- Acknowledge the **source** of the data
- Mention the date the reuse was **last updated**
- **Not alter the meaning** of the data
- Comply with **GDPR**



data.gouv.fr/licences

**What are we
trying to do?**

What are we trying to do?

 **Display good quality data**

 **Show what can be done thanks to open data**

 **Inspire producers and others reusers**

 **Measure the impact of open data**

 **Make it easy to find what you need**

**what
challenges?**

Challenges on reuses

Regarding publication

- **Curation challenge**
- **Monitoring challenge**
- **Discoverability challenge**

Regarding publishers

- Mostly **civil society**
- Not always **maintained**
- Issue for **impact measurement** and advocacy

**How we are
doing it?**

REUSERS

DATA DISCOVERABILITY

Portal features



Tags



Certified organisations



Recommended datasets

The screenshot shows a profile page for 'Santé publique France'. At the top, a blue banner reads 'Service public certifié' and 'L'identité de ce service public est certifiée par Etalab'. To the right is a red circular stamp with a checkmark and the text 'PRODUCTEUR CERTIFIÉ'. Below the banner, the organization's logo consists of a network of colored dots (blue, red, and white) followed by the text 'Santé publique France'. A detailed description below the logo states: 'Santé publique France est l'agence nationale de santé publique. Crée en mai 2016 par ordonnance et décret, c'est un établissement public administratif sous tutelle du ministère...'. At the bottom, there are three orange buttons with white text: 'VOIR LE PROFIL', 'CONTACTER', and 'SUIVRE'.

Curation & editorial



Curation

- Editorialisation and curation work on data.gouv.fr
- Inventory pages

MEILLEURES RÉUTILISATIONS

COVID-19

INFORMATIONS OFFICIELLES

QUESTIONS/RÉPONSES

Données au 21/04/2020

COVID-19 - France

117 324 cas confirmés

20 796 cumul des décès

Données hospitalières

ETALAB TABLEAU DE BORD DE SUIVI DE L'ÉPIDÉMIE DE CORONAVIRUS EN FRANCE

DERNIÈRES RÉUTILISATIONS

Wikipédia

Données réutilisées dans les articles Wikipédia consacrées aux élections législatives

Visualisation 1 ★ 1 2020 WMF

Pizza jambon fromage

Impact environnemental "PEP"

0.43

Impact par étapes du cycle de vie

Impact par ingrédients

Explorer la base Agribalyse pour découvrir l'impact environnemental de votre assiette

Cette application permet d'explorer la base Agribalyse (ADEME, INRAE), et de visualiser rapidement pour l'un des 2500 produits de la base (même taxonomie que la

Application 3 ★ 0 2020 ADEME

Des zones interdépartementales et frontalières sont encore mal ou peu couvertes par la 4G mobile

Method : La couverture 4G de chaque opérateur (Bouygues Telecom, Free Mobile, Orange et SFR) est analysée à partir d'une grille composée de centaines de points de mesure et de calculs effectués par l'opérateur. Le résultat est la visualisation de la couverture 4G.

Dispositif de mesure mobile point d'unité à couvrir

Prévisions d'au moins 10 ans

Nombreux cas de mal couverts par la 4G mobile en 2016

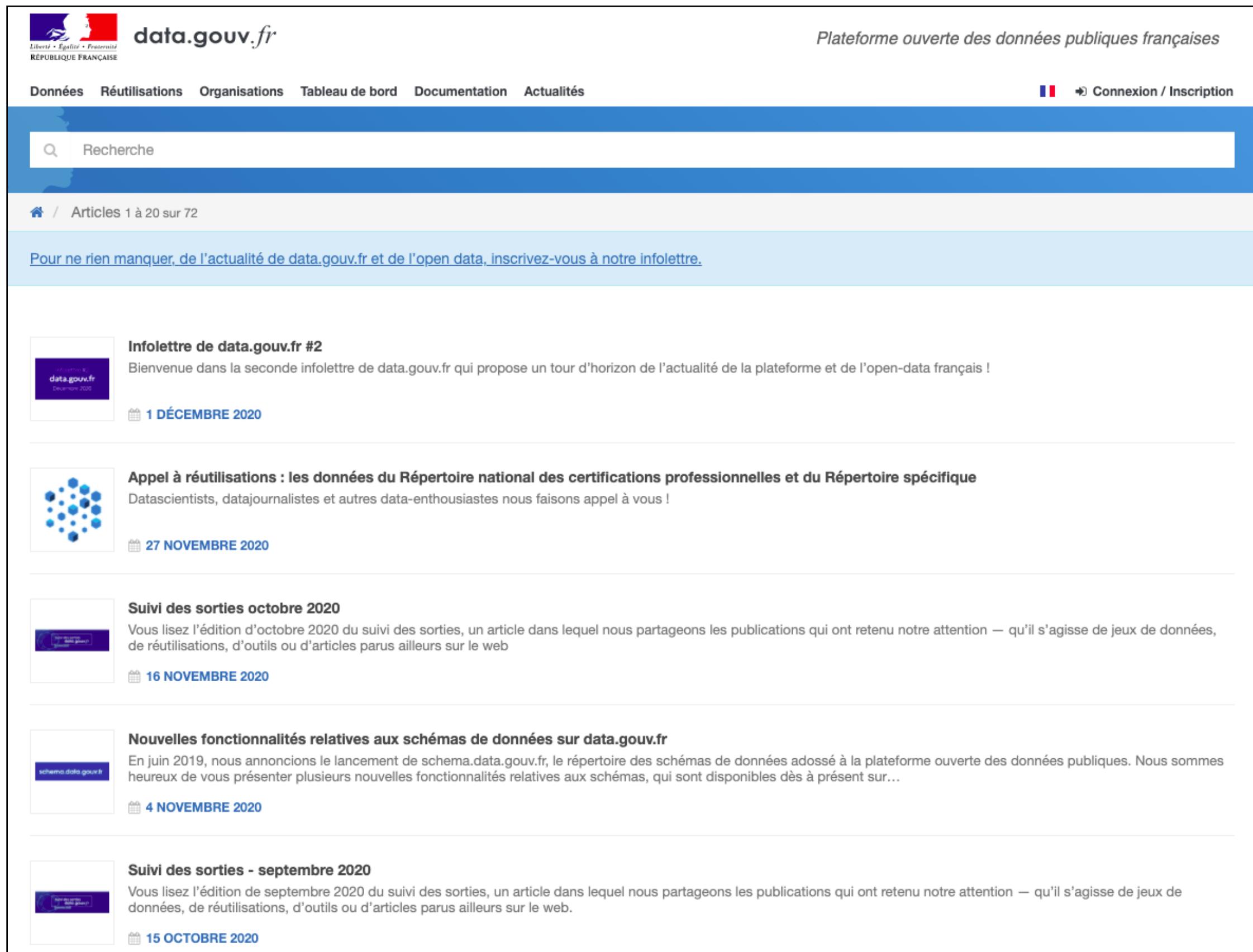
Action opérateur prévue

New Deal Mobile : deux années de progression de la couverture en réseau mobile 4G dans le Grand Est

En analysant les données publiées par l'ARCEP, cette note propose d'observer l'évolution de la couverture mobile 4G dans le Grand Est entre le premier trimestre

Papier 1 ★ 0 2020 Préfecture de la région

Curation & editorial



The screenshot shows the homepage of data.gouv.fr. At the top, there's a navigation bar with links for "Données", "Réutilisations", "Organisations", "Tableau de bord", "Documentation", and "Actualités". Below the navigation is a search bar labeled "Recherche". A banner below the search bar encourages users to subscribe to the newsletter: "Pour ne rien manquer, de l'actualité de data.gouv.fr et de l'open data, inscrivez-vous à notre infolettre.". The main content area displays five news items:

- Infolettre de data.gouv #2** (published 1 DÉCEMBRE 2020) - A welcome message for the second newsletter, featuring a thumbnail of the newsletter logo.
- Appel à réutilisations : les données du Répertoire national des certifications professionnelles et du Répertoire spécifique** (published 27 NOVEMBRE 2020) - An article calling for reuse of specific datasets.
- Suivi des sorties octobre 2020** (published 16 NOVEMBRE 2020) - A summary of notable publications from October 2020.
- Nouvelles fonctionnalités relatives aux schémas de données sur data.gouv.fr** (published 4 NOVEMBRE 2020) - An announcement about new schema-related features.
- Suivi des sorties - septembre 2020** (published 15 OCTOBRE 2020) - A summary of notable publications from September 2020.

Content

- Monthly article to showcase the most important publications
- Specific article on high potential datasets
- Call to reusers

REUSERS

CURATION

Monitor the publications

- Daily watch by the team
- Direct contact with reusers
- Moderation dashboard

The screenshot shows the moderation dashboard for data.gouv.fr. At the top, there's a logo for the French Republic (Liberté, Égalité, Fraternité) and the text "data.gouv.fr". Below the logo, there are links for "Jeux de données", "Réutilisations", and "Organisations". On the right, there are buttons for "se connecter" and a search icon. The main heading is "Bienvenue sur le tableau de bord de modération de data.gouv.fr". Below the heading, there's a button labeled "connectez-vous". A large section displays various statistics in rounded rectangles:

- 35999 datasets
- 7604 discussions
- 24111 followers
- 130 max_dataset_followers
- 133 max_dataset_reuses
- 1017 max_org_datasets
- 572 max_org_followers
- 114 max_org_reuses
- 74 max_reuse_datasets
- 324 max_reuse_followers
- 2651 organizations
- 0 public-service
- 195212 resources
- 2446 reuses
- 65446 users

**Responding to
curious visitors'
needs**

Who are the curious visitors?

- A lot of people are just looking for information on data.gouv.fr
- Not our first mission

```
-- Top 11 réutilisations mois dernier (visites) --
743585 carte-qui-calcule-le-rayon-de-20-km-autour-de-chez-moi-pour-
le-confinement
633495 confinement-comment-mesurer-1-km-autour-de-chez-soi-autour-
du-domicile-application-pour-calculer-un-rayon-1-km-autour-du-
domicile
199650 carte-du-rayon-de-1-km-autour-de-chez-moi-pour-le-
confinement
31555 covid-19-carte-avec-rayon-de-20-km-vol-doiseau-autour-de-
votre-adresse
26496 covid-tracker-suivez-lepidemie-du-covid19-en-france-et-dans-
le-monde
15985 comment-mesurer-20-km-autour-de-son-domicile-de-chez-soi-
application-pour-calculer-un-rayon-de-confinement-20-km-autour-de-
sa-maison
12505 statistiques-sur-la-pandemie-de-coronavirus-covid-19-
rapportees-au-nombre-dhabitants-par-pays
12131 carte-de-la-zone-de-sortie-de-1-km-a-vol-doiseau-autour-du-
domicile-pour-le-reconfinement
12085 reuses
9597 tableau-de-bord-de-suivi-de-lepidemie-de-coronavirus-en-france
```

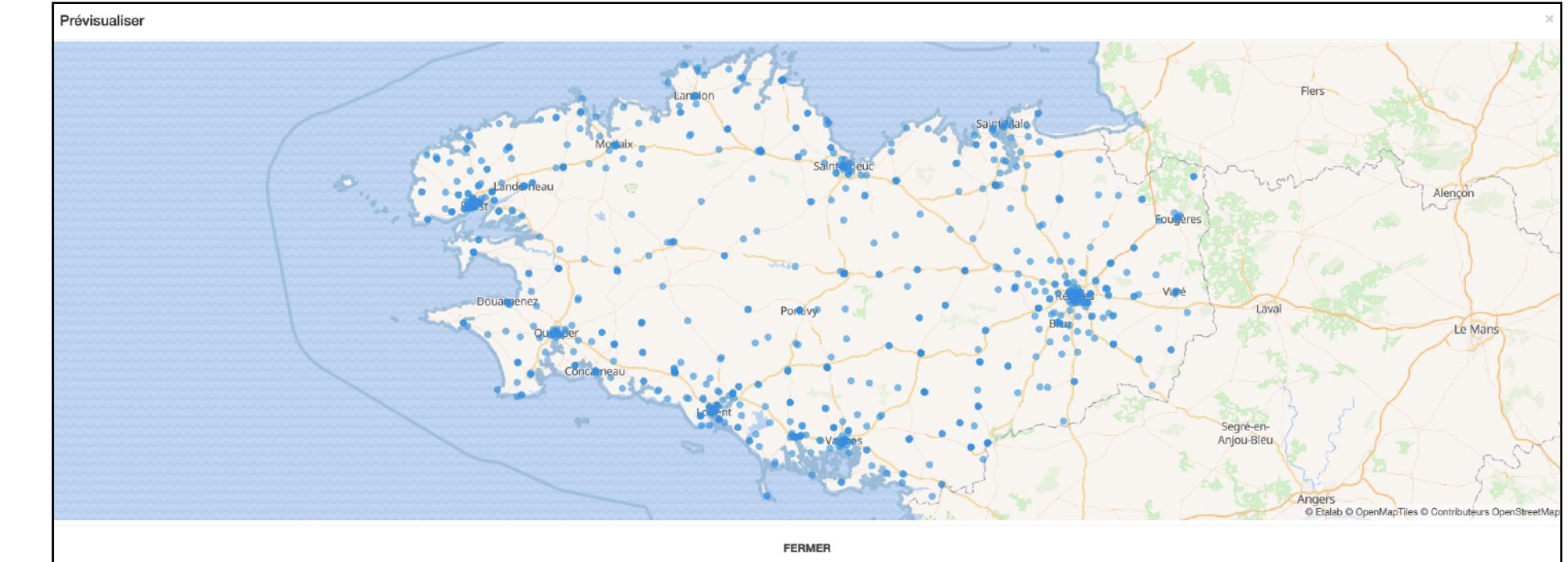
VISITORS

DATA VS INFORMATION

Easy access to information

Editorial work but also
some portal features

- Preview tool
- Recommended reuses



Réutilisations recommandées



Moteur de recherche des personnes décédées (matchID)

Moteur de recherche des décès, gratuit (limité à une requête par seconde) et opensource. Basé sur les fichiers des personnes décédées INSEE -

[Application](#) 1 2 2020

Jeux de données recommandés



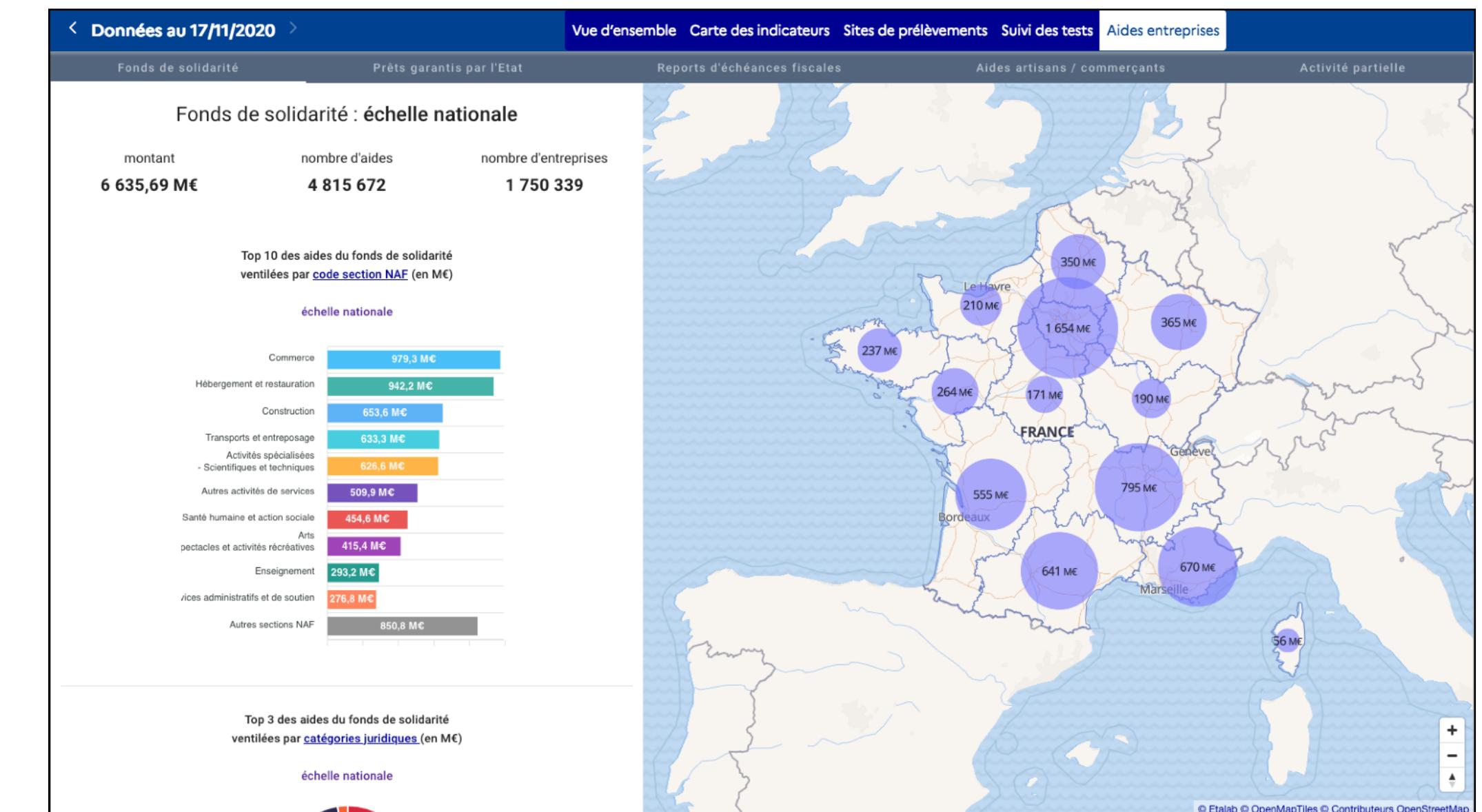
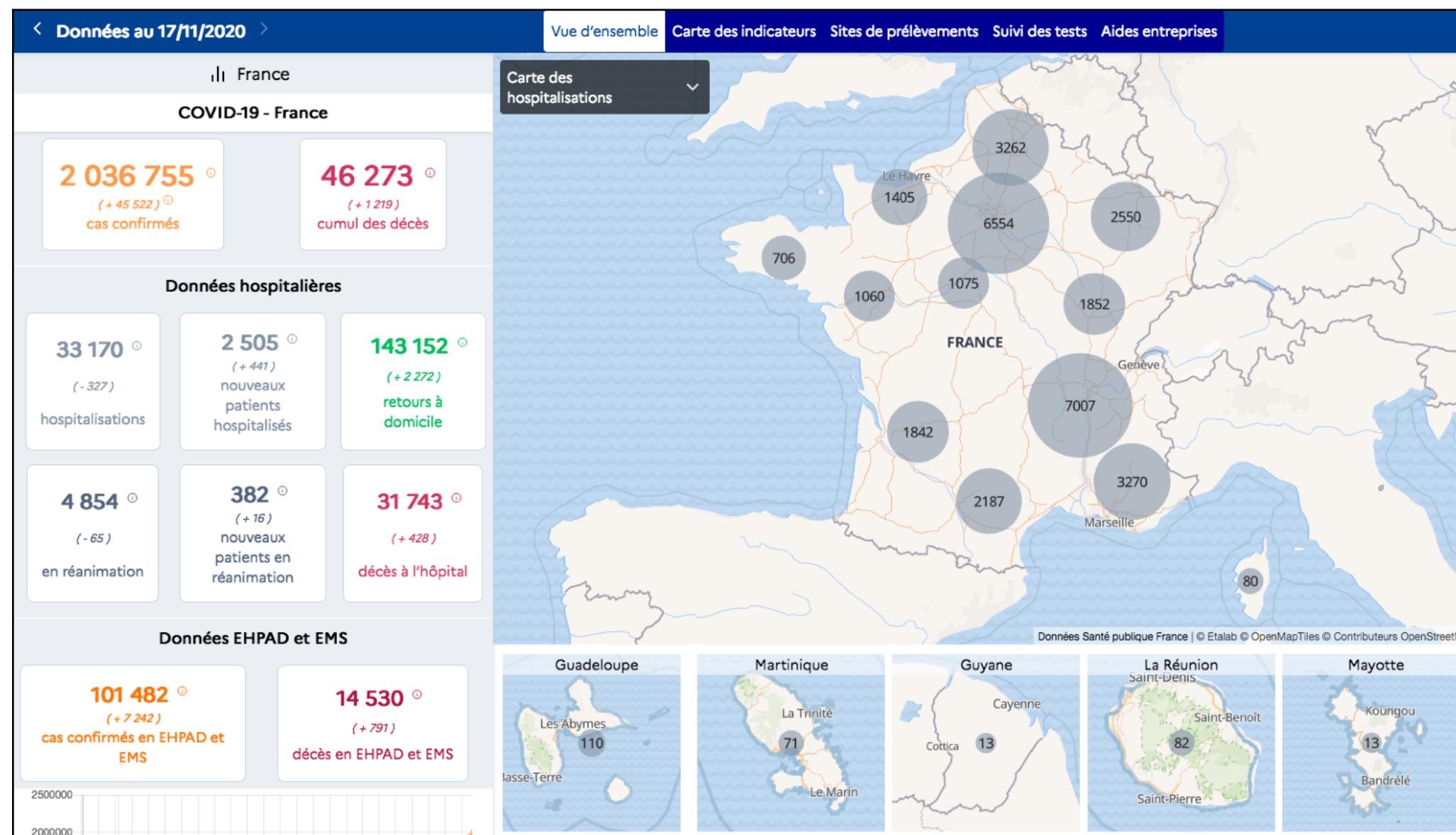
Population mariages deces naissances

Les données publiées comptabilisent l'ensemble des mariages célébrés depuis 1891. Les données sont classées par année, avec le numéro de l'acte

3 0 0 Rodez Commune française

COVID-19

Make data discoverable and understandable



Monitoring the epidemic

Monitoring support to companies



gouvernement.fr/info-coronavirus/carte-et-donnees

**Build a
community**

COMMUNITY

Portal that fosters exchanges

Community ressources

 Contributions communautaires
Explorer avec data.gouv.fr

RESSOURCES COMMUNAUTAIRES

 Répertoire SIRENE au format graphe RDF
Dans le cadre du projet sireneLD, j'ai développé des scripts de transformation du répertoire SIRENE et d'autres données INSEE vers le format RDF, qui permet d'exploiter les...
[n-triples \(7.0Go\)](#) [10](#) [TÉLÉCHARGER](#) 

 Version géocodée, découpée par départements
[csv](#) [54](#) [TÉLÉCHARGER](#) 

 Traitement FME - API SIRENE
Ce traitement, sous le logiciel FME, permet de paramétriser* la requête envoyée à l'API SIRENE de l'Insee et de récupérer les données dans un fichier CSV. Ce traitement est...
[zip \(23.0Ko\)](#) [8](#) [TÉLÉCHARGER](#) 

Discussions

DISCUSSIONS [TRIER PAR ▾](#)

« < 1 2 3 > »

Modification

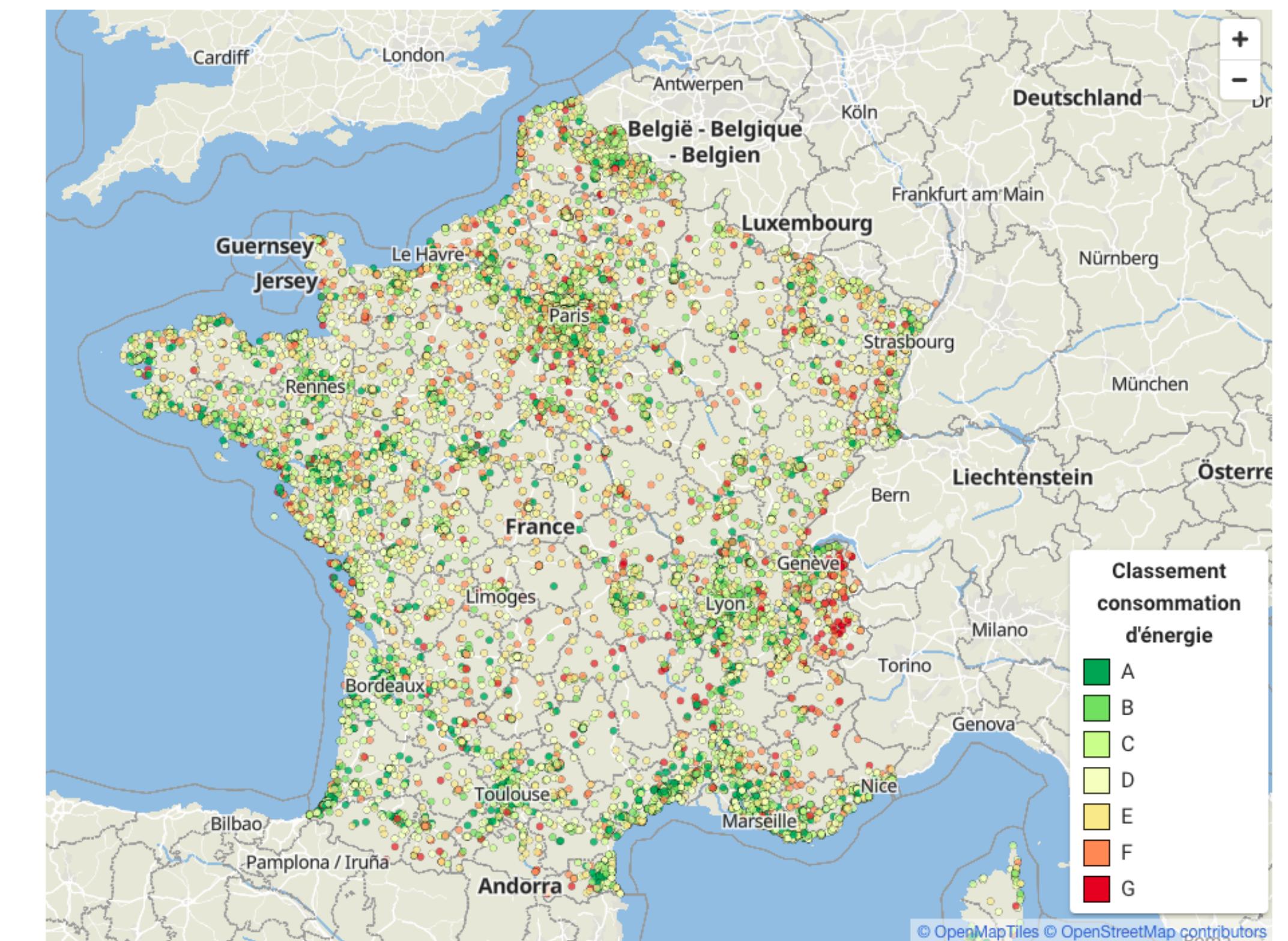
 Marylou Poirier 10 décembre 2020 
Bonjour, a priori il est affiché sur mon dossier : Cette entreprise a exercé son droit d'opposition auprès de l'INSEE. Ses données ne peuvent pas être diffusées publiquement. Cette opposition empêche l'Insee de nous communiquer vos informations et vous bloque pour faire l'affectation dans Datadock.
Or j'ai besoin de faire mon dossier DATADOCK, donc comment puis je modifier cela pour que mes données deviennent publiques ? Merci !

 Diffusion Sirene Insee 11 décembre 2020 
Bonjour,
voici ci-dessous l'adresse du site adéquat pour effectuer votre demande:
Un service permet de modifier le statut de diffusion de manière dématérialisée:
<https://statut-diffusion-sirene.insee.fr/>
La mise à jour au répertoire Sirene sera effective dès le lendemain de votre demande.
Bien cordialement,

[AJOUTER UN COMMENTAIRE](#)

Community animation

- Foster **exchanges** and promote opening and reuses
- DataFin hackathon on local government financial data held the French Senate
- RenovAction hackathon dedicated to energy-efficient renovation of buildings



MERCI !

 mathilde.hoang@data.gouv.fr

 antonin.garrone@data.gouv.fr

 mario.restuccia@data.gouv.fr

RESOURCES

 _DINUM

 etalab

 datagouvfr

 etalab

 etalab

 numerique.gouv.fr

 etalab.gouv.fr

 data.gouv.fr