

Cartographie Collaborative des Données de l'État

La Cartographie Collaborative des Données de l'État est un répertoire de métadonnées de l'État Français. Elle décrit les données gérées par les administrations Françaises, leur organisation et leur gouvernance.

Objectifs et usages

La Cartographie Collaborative des Données de l'État a pour objectif:

- D'identifier les référentiels nationaux et de les valoriser.
- De modéliser la gouvernance des données pour rationaliser leur accès, disponibilité et sécurité et ainsi faciliter leur circulation entre les administrations.
- D'évaluer la qualité globale des données gérées par l'État.
- De généraliser les politiques fondées sur les données.

Composition de la cartographie

La Cartographie Collaborative des Données de l'État est une base de données proposant 3 niveaux:

1. Un annuaire des gestionnaires de données de l'État, issu de l'annuaire de l'administration mis à jour par la DILA.
2. Une liste à vocation exhaustive de toutes les bases de données de l'État, et de leurs gestionnaire identifié.
3. Pour chaque base, une représentation complète et à jour de sa structure de données.

La cartographie a pour objectif de distribuer des informations structurées, claires et pertinentes sur toutes les données de l'État, qu'elles soient ouvertes ou non. Seuls l'annuaire et la liste des bases sont pour le moment publiés en Open Data sur la plateforme [Data.Gouv.fr](https://data.gouv.fr).

Périmètre de la cartographie

Gestionnaires

Sont concernés par la cartographie des données toutes les entités répertoriées par l'annuaire de l'administration, mis à jour par la [Direction Légale de l'Information Légale et Administrative](#) (DILA). Les administrations territoriales, les [Groupements d'Intérêt Public](#), les établissements publics administratifs (EPA) et les établissements publics industriels et commerciaux (EPIC) sont également dans le périmètre de la cartographie.

Bases de données

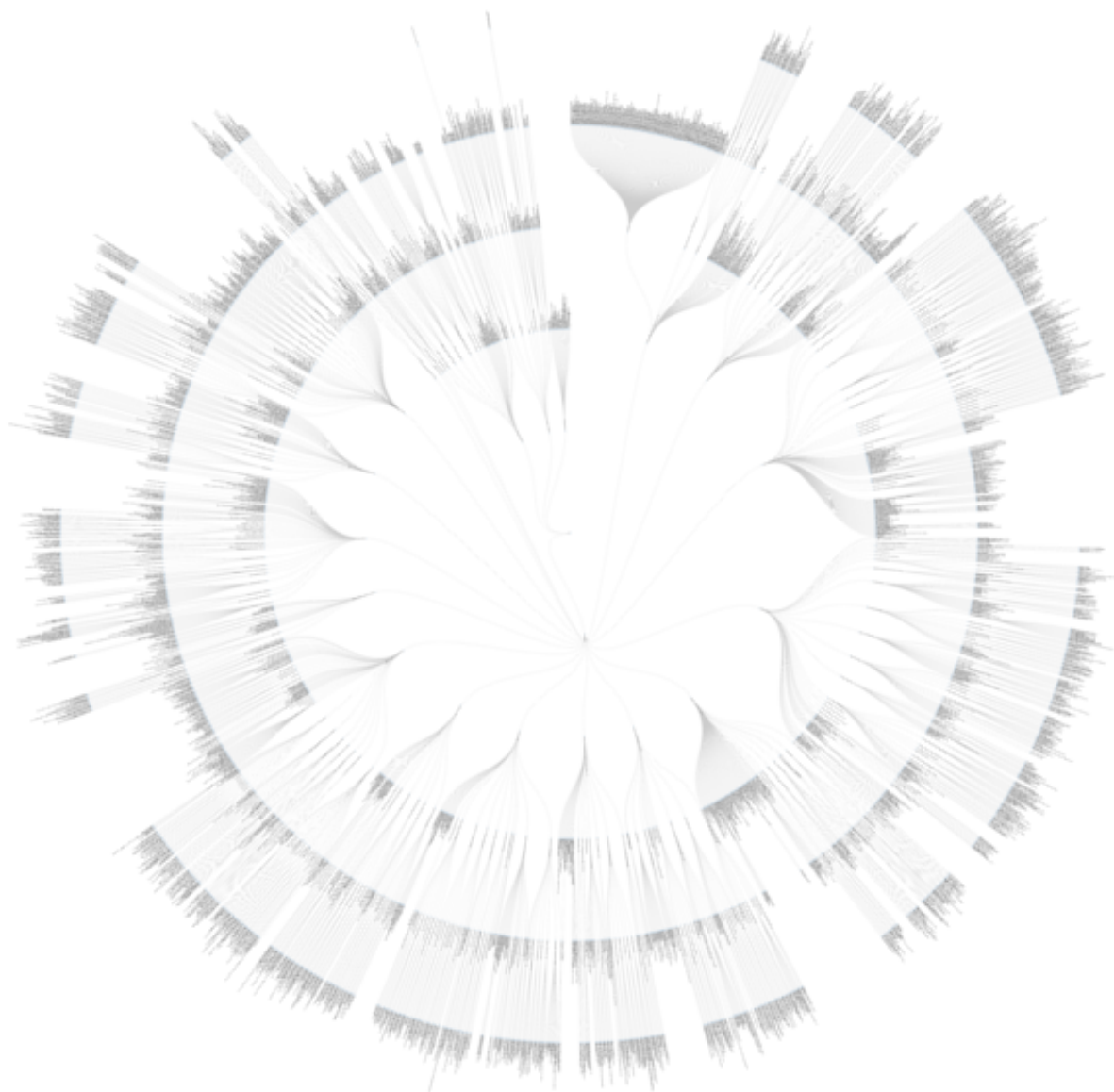
Une base de données est un ensemble de données stockées au même endroit et structurées selon un schéma commun. Elle répond généralement à une problématique et permet à différents acteurs d'accéder ou de mettre à jour les données.

- La cartographie ne répertorie pas pour le moment les données déstructurées.
- Un fichier plat tel qu'un fichier Excel peut être considéré comme une base de données dès lors qu'un processus de mise à jour et d'accès aux données a été clairement établi. En revanche, un fichier Excel ne constitue pas un référentiel tel que défini par le [Cadre Commun d'Architecture des Référentiels de données](#)

Annuaire de gestionnaires de données

Les gestionnaires de données de l'État sont les entités répertoriées par l'annuaire de l'administration mis à jour par la [DILA](#). Cet annuaire est consultable via une [interface web en ligne](#) ou sous la forme d'un [jeu de données ouvert publié sur Data.gouv.fr](#). La dernière version de l'annuaire intégrée à la cartographie contient **3799 administrations**.

Chaque entité de l'annuaire est rattachée à son administration parente, l'ensemble des administrations pouvant être représenté comme un graphe non orienté acyclique et connexe dont voici une représentation graphique:



Restent à intégrer:

- Les administrations territoriales de l'annuaire des services publics.
- Les [Groupements d'Intérêt Public](#), dont la liste est [publiée sur Legifrance](#).
- Les Établissements publics: [EPA](#) et [EPIC](#)

L'historisation de ces annuaires est prévue.

Liste des bases de données

La cartographie décrit [406 bases de données](#) pour [96 gestionnaires](#). Parmi ceux-ci, [42 gestionnaires](#) ne sont pas encore identifiés dans l'annuaire des administrations.

Sources

La liste des bases de données et leurs attributs est mise à jour automatiquement ainsi que de manière collaborative. Des [flux](#) ont été développés pour intégrer différentes sources telles que les cartographies thématiques publiées en Open Data.

- [La cartographie des bases de données publiques en santé.](#)
- [La cartographie des données publiques du logement](#)

Pour l'enrichissement ponctuel de la cartographie, la modification du fichier [bases.json](#) suivi d'une [pull request](#) sur le [répertoire Git de la cartographie](#) est pour le moment nécessaire. Ce fichier est [régulièrement intégré](#) au référentiel de la cartographie.

En raison de la diversité des sources, des données et informations incomplètes, incorrectes ou impertinentes sont susceptibles d'exister.

Modèle de données

Un modèle de données semi-structuré permet de décrire chaque base de données. Ce modèle évolue au fur et à mesure des mises à jour et de l'enrichissement de la cartographie. de nouveaux champs sont donc susceptibles d'apparaître au cours du temps. Seuls les champs **nom** et **gestionnaire** sont indispensables pour décrire l'existence d'une base de données au sein de cette cartographie.

La liste des champs de la liste des bases de données est volontairement limitée, tandis que

Sont décrits ci-dessous tous les champs actuellement susceptibles d'être renseignés pour une base de données de la cartographie.

Nom Le nom complet de la base de données, tel qu'il apparaît dans les textes officiels. Ce champ est obligatoire et renseigné pour toutes les bases de données référencées dans la cartographie.

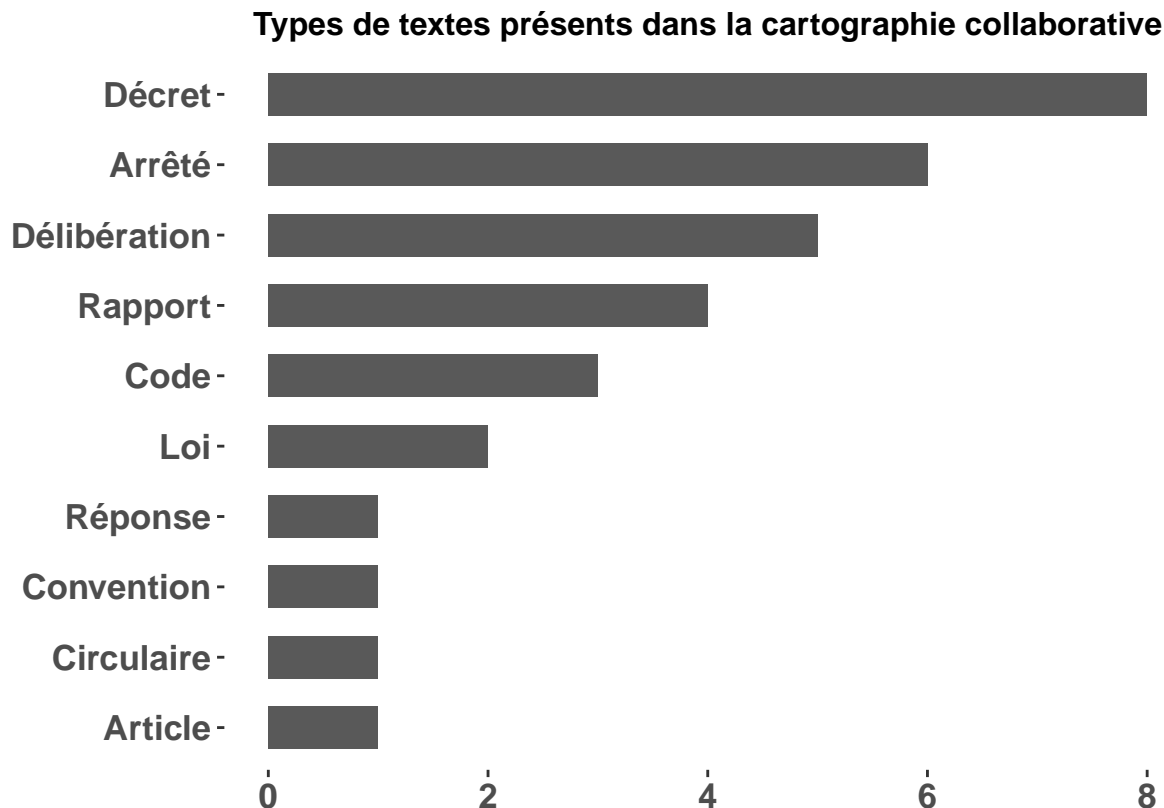
Gestionnaire Le nom du gestionnaire de la base de données, tel qu'il apparaît dans l'[annuaire des services publics](#). Comme le nom complet de la base de données, ce champ est obligatoire. Les champs **gestionnaire** et **nom** forment à eux deux la [clé composite](#) permettant d'identifier chaque base de données de la cartographie.

De nombreuses bases de données sont co-gérées par plusieurs administrations. Dans ce cas, le champ **gestionnaire** devient une liste contenant les tous les co-gestionnaires.

Acronyme L'acronyme de la base de données, ou nom informel de la base, lorsqu'il existe. Ce champ n'est pas obligatoire. Il est actuellement renseigné pour 55 bases de données, soit 13.55% des bases de la cartographie.

Description La description est un champ texte libre décrivant la base de données. Sa taille n'est pas limitée et cette description vise à être la plus complète et exhaustive possible.

Textes Le champ **textes** est une liste des textes officiels publiés par des administrations faisant référence à la base de données. Différents types de textes sont référencés dans la cartographie. En voici la distribution:



Alimentation Les flux entre bases de données sont renseignés dans le champ **alimentation**. Il comporte la liste des organismes alimentant la base de données, ou directement les bases d'origine des données. Le flux de données n'est pas décrit à ce niveau, simplement référencé.

URL Le champ **url** ou [Uniform Resource Locator](#) contient une adresse web vers le portail officiel de la base de données, si il existe. Il s'agit de sites web permettant de consulter le contenu des bases par le biais d'un moteur de recherche lorsque celles ci sont ouvertes.

Références Les adresses web évoquant les bases de données de la cartographie sont listées dans le champ **références**. Contrairement au champ **url**, il ne s'agit pas du portail officiel de la base.

Commentaire Le champ **commentaire** est un champ libre, tout comme le champ **description**. Il regroupe toutes les informations n'ayant pas leur place dans la description de la base de données.

Datasets Le portail Open Data de l'État [data.gouv.fr](#) recense les datasets, ou jeux de données ouverts de l'administration. Les gestionnaires, ici de datasets et pas de bases de données, y sont appelés "producteurs". La cartographie collaborative des données permet de retracer la base d'origine des datasets sur le portail, lorsque le champ **datasets** de la cartographie est renseigné pour une base de données, et que la ou les informations que le champ contient sont conformes.

Le champ **datasets** contient les "slugs" des jeux de données de [data.gouv.fr](#). Il s'agit de la dernière partie de l'url du jeu, pouvant être trouvée en haut du navigateur. Le "slug" est encadré en rouge sur cet aperçu du portail [data.gouv.fr](#):

https://www.data.gouv.fr/fr/datasets/balances-comptables-des-communes

data.gouv.fr

Plateforme ouverte des données publiques françaises

Découvrez l'OpenData Données Tableau de bord Événements Etalab CADA

Recherche Thématiques CONTRIBUTEZ !

Balances comptables des communes


Ce jeu de données provient d'un service public certifié

Balances 2013 des budgets principaux et budgets annexes des communes Balances 2014 des budgets principaux et budgets annexes des communes

Ressources

| ZIP | Balances comptables 2013 des communes |
|-----|---|
| | Dernière modification le mardi 17 novembre 2015 |

Producteur



Le ministère des Finances et des Comptes publics prépare et met en œuvre la politique du Gouvernement en matière

401 producteurs de données publient actuellement 21785 datasets sur [data.gouv.fr](https://www.data.gouv.fr). 40 producteurs du portail sont présents dans l'annuaire des administrations, soit 9.98%.

15 datasets sont renseignés dans la cartographie. Il est donc possible de retracer la base de données d'origine de 0.07% des jeux de données publiés sur [data.gouv.fr](https://www.data.gouv.fr).

Wikidata Wikidata est un projet de base de données libre éditée de manière collaborative et hébergée par la [Wikimedia Foundation](https://www.wikimedia.org/). Il s'agit tout comme la partie *Bases de données* de la cartographie d'une base collaborative structurée, toutefois à visée plus exhaustive que la cartographie.

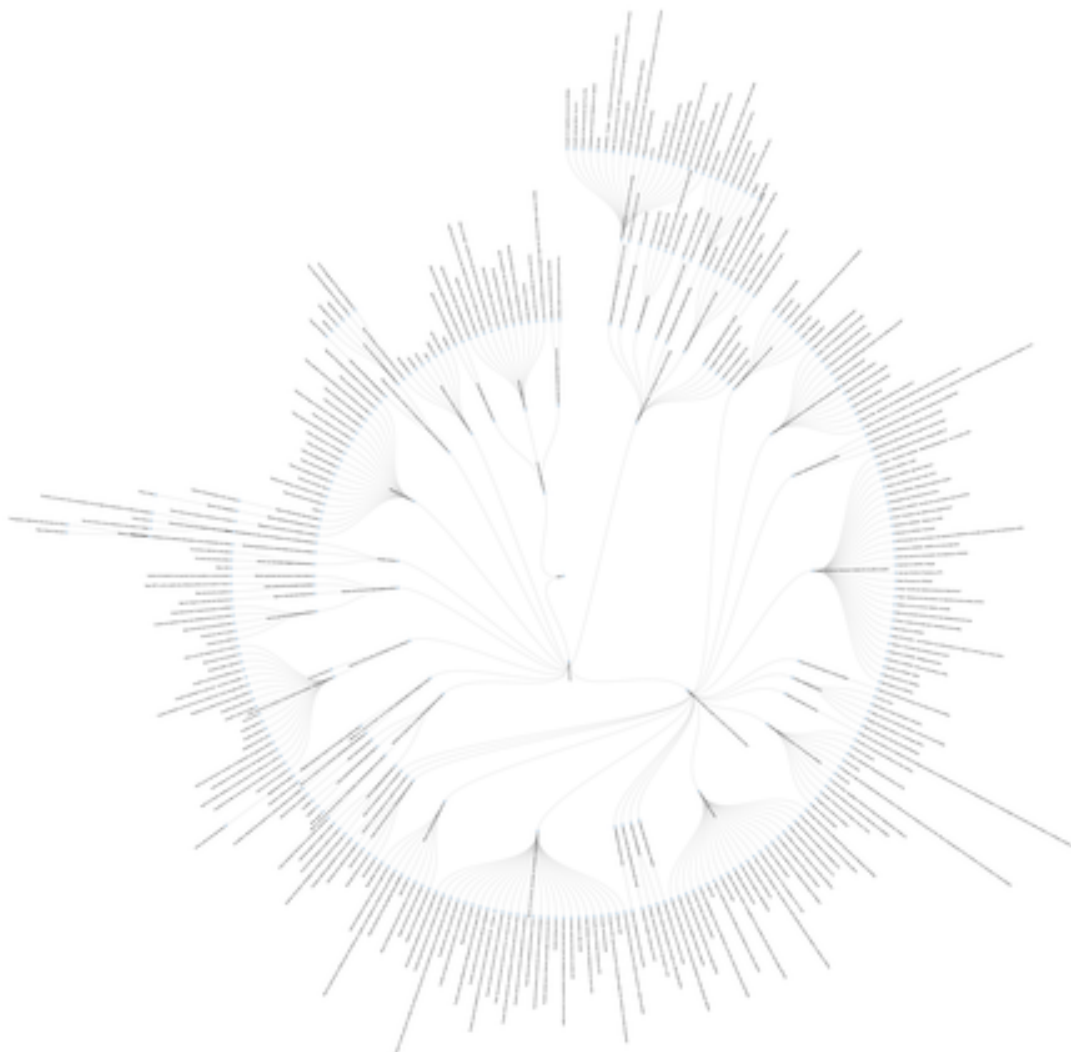
Les identifiants des bases décrites dans Wikidata sont renseignées dans le champ `wikidata` pour permettre une interopérabilité entre les 2 projets.

Exemple: La Base Adresse Nationale Les données de la cartographie sont stockées sous forme de [JSON](#). Voici comment la Base Adresse Nationale est décrite:

```
"Base Adresse Nationale": {
  "description": "Pour que les services d'urgence arrivent au bon endroit, pour vous permettre de réa",
  "textes": {
    "Convention BAN": "http://openstreetmap.fr/f/Convention%20BAN%2015%20avril%202015%20SCAN%20SIGNE.",
  },
  "gestionnaire": [
    "Mission Etalab",
    "Institut national de l'information géographique et forestière",
    "Groupe La Poste"
  ],
  "liens": "https://adresse.data.gouv.fr/",
  "alimentation": [
    "Code Géographique Officiel",
    "Groupe La Poste",
    "Direction générale des finances publiques"
  ],
  "datasets": "ban-base-adresse-nationale",
  "acronyme": "BAN"
}
```

Représenter la cartographie

Considérées comme entités dépendantes de leur gestionnaires, les bases de données peuvent être intégrées à l'arbre des administrations. La visualisation est alors similaire à celle de l'annuaire:



Répertoire des structures de données

La documentation des structures de données des bases référencées ne sont pas encore à l'ordre du jour. Des expérimentations sont à venir, notamment sur quelques bases ouvertes tel que la [Base Adresse Nationale](#) et la présente cartographie.

Structure technique de la cartographie

La cartographie collaborative est une base de données, à l'image des entités qu'elle référence. À ce titre, elle est présente dans son propre référentiel, modélisée en suivant le même schéma que toutes les autres bases. Bien que la cartographie des bases de données soit pour l'instant le seul dataset produit, la cartographie intègre d'autres sources.

- [L'annuaire des services publics](#) publié par la [DILA](#).
- [La liste de des jeux de données publiés sur le portail Data.gouv.fr](#)

La cartographie dispose de copies de ces données le plus à jour possible, dont les versions font uniquement référence dans le cadre local de la cartographie. Ces versions ne sont pas publiées et les originaux doivent être la source de toute autre réutilisation.

Portail collaboratif

Le déploiement d'un portail collaboratif est prévu pour permettre à chacun de contribuer à la partie *Bases de données* de la cartographie sans avoir à utiliser [Git](#).

Bibliographie

- [Les apports de la méthode MDM dans la performance du SI des entreprises.](#)
- [Data Management : qualité des données et compétitivité](#)
- [MDM : Enjeux et méthodes de la gestion des données](#)
- [Defining Master Data](#)
- [Enjeux business des données](#)
- [We need open and vendor-neutral metadata services](#)
- [Big data at a crossroads: Time to go meta \(on use\)](#)
- [Michel Volle, De l'Informatique, savoir vivre avec l'automate](#)
- [Mapping Big Data, A Data-Driven Market Report](#)
- [Enjeux du Master Data Management](#)