

API Données locales

9 juillet 2021

Données locales



Table des matières

1	Présentation du service Web	2
1.1	Présentation des données accessibles	3
2	Paramètres en entrée	4
2.1	Jeux de données (jeu_donnees)	4
2.2	Croisements	4
2.3	Modalités	5
2.4	Niveaux géographiques (nivgeo)	5
2.5	Codes géographiques (codgeo)	6
2.6	Remarques générales	6
3	Paramètres de sortie	7
4	Documentation associée	10
4.1	Détail des fichiers Excel de documentation par source	10
4.2	Package R	12
5	Annexes	13
5.1	Codes article des libellés géographiques	13
5.2	Cas d'utilisation du package R	13

1 Présentation du service Web

L'API 'Données locales' permet d'accéder aux données localisées à la commune, diffusées sur insee.fr dans la rubrique 'chiffres détaillés'¹, sous forme de cubes prédéfinis.

Les cubes prédéfinis sont ceux utilisés pour l'élaboration des tableaux et graphiques en ligne, correspondant aux sources suivantes : recensement de la population, état civil, répertoire des entreprises et des établissements, fichier localisé social et fiscal, fichier localisé des rémunérations et de l'emploi salarié et offre d'hébergement touristique.

Les statistiques sont disponibles pour tous les zonages géographiques allant de la commune à la France entière, c'est-à-dire pour les zonages administratifs (communes, arrondissements municipaux, arrondissements, intercommunalités/EPCI, départements, régions et France), ainsi que pour les zonages d'étude (zones d'emploi, aires d'attraction des villes, unités urbaines).

La recherche s'effectue à partir d'une source, d'un croisement de variables et d'un code géographique.

La consultation des résultats du recensement de la population (RP) étant destinée à un large public, la diffusion de certains tableaux est restreinte pour en garantir une interprétation de qualité. Ainsi la même finesse de croisement n'est pas disponible pour toutes les zones géographiques. Le détail des restrictions est précisé dans la documentation du recensement (*cf.* groupes de diffusion).

Exemple : certains tableaux ne sont élaborés que pour les zones géographiques de plus de 2000 habitants.

Nota Bene : Les bases de données du RP mises à disposition sur insee.fr ne sont pas soumises à ces restrictions, notamment pour permettre un usage expert des données et la reconstitution éventuelle d'information sur des regroupements de communes à façon.

1. 'chiffres-clés' pour les populations légales

1.1 Présentation des données accessibles

Nom court de la source de données	Nom détaillé de la source	Millésimes disponibles
RP	Recensement de la population	1999, de 2006 à 2018
BDCOM	Séries historiques du recensement de la population (depuis 1968)	de 2012 à 2018
Popleg	Populations légales (issue du recensement de la population)	de 2006 à 2018
RFD	Données de l'état-civil (naissances et décès)	de 2007 à 2019
REE	Répertoire des entreprises et des établissements (issu de Sirene)	de 2009 à 2020
FILOSOFI	Fichier localisé social et fiscal	2017 et 2018
Flores	Fichier localisé des rémunérations et de l'emploi salarié	2017 et 2018
TOUR	Tourisme (offre d'hébergement)	de 2013 à 2021

Pour afficher les données dans l'API, il faut renseigner un code **jeu de données** (indiquant la source et le millésime des données), une sélection de variables (**croisement**) et les **modalités** que l'on souhaite visualiser, ainsi que le **niveau** et le **code géographique** souhaité. Le cube formé par ces différents paramètres est ensuite restitué par l'API.

Paramètres en entrée de l'API	Description
jeu_donnees	Code du jeu de données (cf. documents par source) : il se compose soit d'un nom court de la source + millésime (RP2015), soit géographie de projection + nom court de la source + millésime (GEO2019RP2016)
croisement	Combinaison de variables souhaitées sous la forme var1-var2-... (exemple sexe-age)
modalite	Valeurs des variables souhaitées séparées par un .
nivgeo	Niveau géographique
codgeo	Code géographique

Pour utiliser l'API directement sans passer par le catalogue des API, l'url se construit de la manière suivant :

https://api.insee.fr/donnees-locales/V0.1/donnees/geo-croisement@jeu_donnees/nivgeo-codgeo.modalite

Les données sont disponibles pour les zonages administratifs (communes, arrondissements municipaux, arrondissements, EPCI, départements, régions et France), ainsi que pour les zonages d'étude (zones d'emploi, aires d'attraction des villes, unités urbaines). Pour certains croisements, les données ne sont pas diffusées sur tous les codes géographiques, pour garantir une interprétation de qualité. Ainsi, l'offre de données est plus restrictive que dans l'offre des bases mises

à disposition sur insee.fr. Le détail des restrictions est précisé dans la documentation Excel (*cf.* groupes de diffusion).

Les **données renvoyées** sont au format XML ou json. Elles se décomposent en 4 parties distinctes contenant :

- des informations sur la source demandée (code, libellé, millésime, filtre géographique) ;
- des informations sur la ou les zone(s) géographique(s) demandée(s) ;
- les différentes modalités de chaque variable demandée ;
- la valeur associée.

Pour chaque source, un fichier Excel documentant toute l'information présente dans l'API est mis à disposition. Le contenu de ce fichier est détaillé par la suite.

2 Paramètres en entrée

2.1 Jeux de données (jeu_donnees)

Le code jeu de données se compose du nom de la source, du millésime des données et parfois du millésime géographique de diffusion.

Avant avril 2019, les données locales sont diffusées dans la géographie de production des données. Le nom du jeu de donnée se composait uniquement du nom de la source et du millésime des données (RP2015 par exemple). Le millésime géographique des données est indiqué dans la documentation Excel de la source.

À partir d'avril 2019, l'Insee a décidé de diffuser les données locales dans la dernière géographie disponible au moment de la diffusion. Afin de pouvoir déterminer les données qui sont retraitées, le millésime géographique est renseigné dans le nom du jeu de données. Celui-ci aura pour préfixe GEOXXXX avec XXXX le millésime géographique (GEO2019RFD2010 par exemple).

2.2 Croisements

Les données statistiques diffusées se déclinent selon des variables descriptives (sexe, classe d'âge, nombre d'étoiles...). Les données peuvent se calculer selon une variable ou selon un regroupement de variables (selon le sexe et la tranche d'âge par exemple).

La sélection de variables est appelée croisement dans le reste du document. Celle-ci correspond à une combinaison de variables sous la forme var1-var2-...-varN (sexe-age par exemple).

Chaque variable a une liste de codes associés : ce sont les modalités prises par la variable (sexe a pour modalités homme et femme).

Quelques règles :

- l'ordre des variables n'a pas d'importance ;
- l'usage des minuscules ou majuscules est indifférent ;
- les croisements sont définis a priori : si les variables var1 et var2 existent, le croisement var1-var2 n'existe pas nécessairement ;
- les croisements ne sont pas identiques pour tous les millésimes ;
- les croisements ne doivent pas contenir d'espace.

Les combinaisons possibles de variables sont détaillées dans l'onglet **mesure_croisement** des documents par source.

2.3 Modalités

Le paramètre modalité doit contenir les modalités voulues des variables renseignées dans croisement.

Il doit contenir autant d'éléments que de variables utilisées dans le paramètre croisement et dans le même ordre. Par exemple si le croisement est de la forme var1-var2-var3, le paramètre modalité doit être de la forme X.Y.Z avec X,Y et Z pouvant valoir :

- "ALL" si on veut toutes les modalités disponibles de la variable concernée ;
- une modalité précise : par exemple "1" pour la variable "Sexe" qui renverra donc la modalité "Hommes". Beaucoup de variables disposent de la modalité ENS ("Ensemble"), qui correspond au total ;
- plusieurs modalités précises séparées par un "+" : par exemple "1+2" renverra les modalités "Hommes" et "Femmes" de la variable "Sexe".

Les modalités pour chaque variable sont détaillées dans l'onglet **var_modalite** des documents par source.

2.4 Niveaux géographiques (nivgeo)

Les niveaux géographiques suivants sont disponibles :

- COM : communes et arrondissements municipaux,
- DEP : départements,
- REG : régions,
- METRODOM : France métropolitaine,
- FE : France,
- ARR : arrondissements,
- EPCI : intercommunalités,
- AAV2020 : aires d'attraction des villes 2020,
- UU2020 : unités urbaines2020,
- ZE2020 : zones d'emploi 2020,
- AU2010 : aires urbaines 2010,
- UU2010 : unités urbaines2010,
- ZE2010 : zones d'emploi 2010.

Attention : Les zonages d'études ont été revus en 2020. Les zones d'emploi 2020 ont été diffusées le 10 septembre 2020 et remplacent les zones d'emploi 2010. Les aires d'attraction des villes 2020 et les unités urbaines 2020 ont été diffusées le 21 octobre 2020 et remplacent les aires urbaines 2010 et les unités urbaines 2010.

À partir de ces dates, la diffusion des données s'effectuent uniquement sur les nouveaux zonages d'études.

2.5 Codes géographiques (codgeo)

Le code géographique à renseigner est le Code Officiel Géographique pour les zonages administratifs ou le code fourni par l'Insee pour les zonages d'étude. Ces codes peuvent être obtenus sur insee.fr. Le code de la France est "1".

Attention : jusqu'en avril 2019, les données sont diffusées selon le millésime géographique de production. Le recensement de la population 2013 en géographie 2015 ne contient par exemple aucune donnée sur les intercommunalités créées au 1^{er} janvier 2016.

2.6 Remarques générales

Les informations restituées pour un croisement donné peuvent caractériser différentes "populations statistiques" (individu, actif, famille, logement...). Cette notion sera appelée la **mesure**. Celle-ci ne peut pas être renseignée en entrée de l'API, mais elle est précisée dans la sortie grâce à une balise spécifique (<Mesure>).

Chaque jeu de données dispose de croisements et de mesures spécifiques. Ainsi la documentation Excel pour chaque source décrit les différentes combinaisons possibles. Pour chaque ligne, le nom d'un tableau existant sur insee.fr contenant l'information est indiqué. Le tableau peut directement être retrouvé sur insee.fr en indiquant son nom court (POP T1 par exemple) dans le moteur de recherche du site.

Le croisement et la mesure ne sont pas toujours suffisants pour comprendre l'information restituée. La colonne filtre statistique, des documentations Excel par source, permet d'obtenir de l'information supplémentaire qui permet de déterminer les restrictions de champ appliquées à la mesure.

3 Paramètres de sortie

- Une balise <Croisement> qui fournit quelques informations sur la source demandée.
Tout d'abord une information sur le filtre géographique. Attention, le contenu n'est pas toujours fiable. Il faut donc se reporter à la documentation de la source pour ce paramètre. Ensuite, des informations sur la source et son millésime sont disponibles.
- Une ou plusieurs balises <Zone> qui fournissent des informations sur la ou les zone(s) géographique(s) demandée(s).
La zone est déterminée par un libellé sans article (Nccenr) et un code indiquant l'article (tncc). Les modalités du code article sont indiquées en annexe.
- Une ou plusieurs balises <Variable> qui fournissent les différentes modalités de chaque variable demandée.
- Une ou plusieurs balises <Cellule> qui fournissent la valeur associée à une zone, une combinaison de modalités et une mesure donnée.
Par exemple une cellule pourra contenir la valeur de la mesure "Nombre de logements" pour la modalité 1 de la variable tf4 (correspondant aux couples sans enfant) et pour la zone géographique correspondant à la commune de Nantes. Dans l'exemple, le croisement a 2 mesures disponibles. Le nombre de logements pour chaque type de famille est d'abord indiqué, puis la population pour chaque type de famille est indiquée.

Un exemple de réponse est présenté ci-après pour la source RP2015, le croisement TF4 et toutes les mesures :

Paramètres en entrée :

jeu_donnees = RP2015

croisement = TF4

modalités = all

nivgeo = COM

codgeo = 44109

Sortie :

<Donnees>

<Croisement>

<Filtre/>

<Groupe code="FHM_P2000">France hors Mayotte, d'au moins 2 000 habitants</Groupe>

<JeuDonnees code="RP2015">

<Annee>2015</Annee>

<Libelle>Recensement de la population 2015</Libelle>

<Source>Recensement de la population</Source>

</JeuDonnees>

</Croisement>

<Zone codgeo="44109" nivgeo="COM">

<Millesime annee="2017">

<Nccenr>Nantes</Nccenr>

<Tncc>0</Tncc>

</Millesime>

</Zone>

<Variable code="TF4">


```

    <Libelle>Type de famille regroupé (4 postes)</Libelle>
    <Modalite code="ENS" variable="">
      <Libelle>Ensemble</Libelle>
    </Modalite>
    <Modalite code="1" variable="">
      <Libelle>Couple sans enfant</Libelle>
    </Modalite>
    <Modalite code="2" variable="">
      <Libelle>Couple avec enfant(s)</Libelle>
    </Modalite>
    <Modalite code="3" variable="">
      <Libelle>Famille monoparentale composée d'un homme avec enfant(s)</Libelle>
    </Modalite>
    <Modalite code="4" variable="">
      <Libelle>Famille monoparentale composée d'une femme avec enfant(s)</Libelle>
    </Modalite>
  </Variable>
  <Cellule>
    <Zone codgeo="44109" nivgeo="COM"/>
    <Mesure code="NBLOG">Nombre de logements</Mesure>
    <Modalite code="ENS" variable="TF4"/>
    <Valeur>69425.91637</Valeur>
  </Cellule>
  <Cellule>
    <Zone codgeo="44109" nivgeo="COM"/>
    <Mesure code="NBLOG">Nombre de logements</Mesure>
    <Modalite code="1" variable="TF4"/>
    <Valeur>30720.168582</Valeur>
  </Cellule>
  <Cellule>
    <Zone codgeo="44109" nivgeo="COM"/>
    <Mesure code="NBLOG">Nombre de logements</Mesure>
    <Modalite code="2" variable="TF4"/>
    <Valeur>26510.301474</Valeur>
  </Cellule>
  <Cellule>
    <Zone codgeo="44109" nivgeo="COM"/>
    <Mesure code="NBLOG">Nombre de logements</Mesure>
    <Modalite code="3" variable="TF4"/>
    <Valeur>1921.083074</Valeur>
  </Cellule>
  <Cellule>
    <Zone codgeo="44109" nivgeo="COM"/>
    <Mesure code="NBLOG">Nombre de logements</Mesure>
    <Modalite code="4" variable="TF4"/>
    <Valeur>10274.363241</Valeur>
  </Cellule>

```

```

</Cellule>
<Cellule>
  <Zone codgeo="44109" nivgeo="COM"/>
  <Mesure code="POP">Population</Mesure>
  <Modalite code="ENS" variable="TF4"/>
  <Valeur>197897.334421</Valeur>
</Cellule>
<Cellule>
  <Zone codgeo="44109" nivgeo="COM"/>
  <Mesure code="POP">Population</Mesure>
  <Modalite code="1" variable="TF4"/>
  <Valeur>62178.020416</Valeur>
</Cellule>
<Cellule>
  <Zone codgeo="44109" nivgeo="COM"/>
  <Mesure code="POP">Population</Mesure>
  <Modalite code="2" variable="TF4"/>
  <Valeur>103679.837265</Valeur>
</Cellule>
<Cellule>
  <Zone codgeo="44109" nivgeo="COM"/>
  <Mesure code="POP">Population</Mesure>
  <Modalite code="3" variable="TF4"/>
  <Valeur>4854.911575</Valeur>
</Cellule>
<Cellule>
  <Zone codgeo="44109" nivgeo="COM"/>
  <Mesure code="POP">Population</Mesure>
  <Modalite code="4" variable="TF4"/>
  <Valeur>27184.565164</Valeur>
</Cellule>
</Donnees>

```

4 Documentation associée

4.1 Détail des fichiers Excel de documentation par source

Onglet mesure_croisement :

L'information principale est mise à disposition dans cet onglet. Il permet de déterminer pour chaque modalité, les croisements mis à disposition pour chaque jeu de données. Les filtres statistiques et géographiques sont également renseignés. Enfin le nom d'un produit sur insee.fr utilisant le cube est renseigné pour permettre à l'utilisateur d'aller visualiser les données associées.

Description des variables du fichier :

mesure : code de la mesure

croisement : succession de variables séparées par un '-'

filtre_statistique : filtre sur les données qui n'est pas explicité par la mesure ou par le croisement

filtre_geo : filtre appliqué sur les codes géographiques

filtre_geo_avt_2017 : filtre appliqué sur les codes géographiques avant le millésime 2017. Afin d'augmenter l'offre dans l'API, certains filtres géographiques ont été assouplis à partir de la diffusion du recensement 2017. Cette colonne n'est renseignée que pour le recensement de la population.

type_exploitation : type d'exploitation du recensement de la population (principal ou complémentaire). L'échantillon n'étant pas identique dans les deux exploitations, pour un même croisement les données n'auront pas les mêmes valeurs. Cette colonne n'est renseignée que pour le recensement de la population.

nom_tab : nom court du tableau dynamique sur le site insee.fr. Ce nom doit permettre de retrouver le tableau correspondant au croisement sur insee.fr en utilisant le moteur de recherche.

Ensuite, une colonne par jeu de données pour indiquer si le croisement est présent ou non pour les différents millésimes.

Onglet var_modalite :

L'information présente dans cet onglet permet de déterminer les différentes modalités de chaque variable.

Description des variables du fichier :

variable : code variable utilisé dans les croisements

lib_var : libellé de la variable

modalite : code de la modalité

lib_modalite : libellé de la modalité Ensuite une colonne par source pour indiquer si le croisement est présent ou non.

Onglet lib_mesure :

Cet onglet permet d'obtenir le libellé de chaque mesure.

Description des variables du fichier :

mesure : code de la mesure utilisée

lib_mesure : libellé de la mesure

Onglet lib_tableau :

Cet onglet permet de faire le lien entre le nom court du tableau dynamique sur le site insee.fr et son titre long.

Par exemple le tableau POP T0 a pour titre long "Population par grandes tranches d'âges". "

Description des variables du fichier :

nom_tab : nom du tableau sur insee.fr

lib_tab : titre du tableau sur insee.fr

Onglet lib_filtre_geo :

Cet onglet permet d'obtenir le libellé de chaque filtre géographique.

Description des variables du fichier :

filtre_geo : code du filtre géographique

lib_filtre_geo : libellé du filtre géographique

Onglet millésime :

Cet onglet permet de faire le lien entre un jeu de données, le millésime des données et le millésime géographique.

Description des variables du fichier :

jeu_donnees : code du jeu de données

millésime_donnees : millésime des données

millésime_geo : millésime géographique des données diffusées.

Onglet documentation :

Cet onglet contient la documentation relative aux données diffusées pour la source. Des liens vers la documentation sur insee.fr sont présents, ainsi que des avertissements sur certains croisements de certaines années .

Onglet formules :

Cet onglet, disponible uniquement pour le RP, contient les formules permettant de calculer des indicateurs complémentaires disponibles dans les tableaux dynamiques sur Insee.fr.

4.2 Package R

Un package R est mis à disposition sur le compte github d'InseeFrLab.

Le `inseeLocalData` permet d'importer les données présentes dans l'API Données Locales dans une liste contenant 4 objets :

- les données statistiques ;
- les modalités de chaque variable ;
- l'information sur la zone demandée ;
- l'information sur la source et le jeu de données demandé.

Des cas d'utilisation du package sont renseignés en annexe. Ces cas permettent :

- d'obtenir les données pour une requête
- d'obtenir les données pour une série de requêtes
- d'obtenir les données d'une requête pour une liste de codes géographiques

Pour ces 3 opérations, il est possible soit de générer uniquement une liste R, soit de créer des fichiers Excel contenant les données.

5 Annexes

5.1 Codes article des libellés géographiques

code	article	charnière
0		de
1		d'
2	le	du
3	la	de la
4	les	des
5	l'	de l'
6	aux	des
7	las	de las
8	los	de los

5.2 Cas d'utilisation du package R

Installation du package :

```
remotes::install_github("insee-frlab/inseeLocalData")
```

Exemple d'utilisation simple :

Cet exemple permet d'utiliser l'API pour un croisement et un code géographique.

```
croisement <- "NA5_B-ENTR_INDIVIDUELLE"
jeu_donnees <- "GE02017REE2017"
nivgeo <- "COM"
codgeo <- "51108"
modalite <- "all.all"

donneesAPI <- get_dataset(jeton, jeu_donnees, croisement, modalite, nivgeo, codgeo)

donnees <- donneesAPI$donnees # pour accéder aux données
liste_code <- donneesAPI$liste_code # pour accéder aux nomenclatures
info_zone <- donneesAPI$info_zone # pour accéder aux données géographiques
source <- donneesAPI$source # pour accéder à la source
```

Exemple d'utilisation sur plusieurs codes géographiques :

Cet exemple permet d'obtenir un résultat pour un même croisement sur plusieurs codes géographiques. Il est nécessaire dans un premier temps de charger une liste de codes géographiques et leurs niveaux. Dans l'exemple, il s'agit du data.frame liste_code.

```
liste_code <- data.frame(codgeo = c("200023372","74056","74143","74266","74290"),
                        nivgeo = c("EPCI","COM","COM","COM","COM"))
croisement <- "NA5_B-ENTR_INDIVIDUELLE"
jeu_donnees <- "GEO2017REE2017"
modalite <- "all.all"

sortie <- mapply(get_dataset,
                 jeton, jeu_donnees, croisement,
                 modalite, liste_code$nivgeo, liste_code$codgeo, USE.NAMES = TRUE)

donnees <- NULL
info_zone <- NULL
for (i in 1:dim(sortie)[2]){
  donnees <- rbind(donnees, sortie[,i]$donnees)
  info_zone <- rbind(info_zone, sortie[,i]$info_zone)
}

liste_code <- sortie[,1]$liste_code # la liste de code est la même pour tous les codes
géographiques
source <- sortie[,1]$source # la source est la même pour tous les codes géographiques
```

Exemple d'utilisation sur plusieurs croisements :

Cet exemple permet d'utiliser la fonction pour récupérer les données sur plusieurs croisements (sur la même zone géographique ou une zone différente). Les paramètres sont renseignés au préalable dans un data.frame 'fichier', ayant pour variables jeu_donnees, croisement, modalite, nivgeo et codgeo. Le paramètre temporisation est utilisé pour faire une pause de 2 secondes entre chaque requêtes afin de ne pas dépasser le quota de 30 requêtes par minute.

```
fichier <- 'mon fichier'
sortie <- mapply(get_dataset,
                 jeton, fichier$jeu_donnees, fichier$croisement,
                 fichier$modalite, fichier$nivgeo, fichier$codgeo, 2, USE.NAMES = TRUE)

# pour le 1er croisement renseigné dans le fichier en entrée
donnees <- sortie[,1]$donnees # pour accéder aux données du 1er croisement renseigné
dans le fichier
liste_code <- sortie[,1]$liste_code # pour accéder aux nomenclatures du 1er croisement
renseigné dans le fichier
info_zone <- sortie[,1]$info_zone # pour accéder aux données géographiques du 1er
croisement renseigné dans le fichier
source <- sortie[,1]$source # pour accéder à la source du 1er croisement renseigné dans
le fichier
```