



GéoPeuple



# Intégration des données démographiques

## Migration de la base de données Cassini vers GeoPeople

Rapport numéro	L2-3.4
Titre	Intégration des données démographiques
Rédigé par	Christine Plumejeaud (COGIT/IGN), Hervé Le Bras (EHESS)
Etat (en cours / final)	Final
Relu par	Claude Motte – Pascal Cristofoli
Date	22 août 2012 – Relecture finale Octobre 2013

Ce document propose une analyse détaillée de la base de données d'EHESS, qui contient à l'heure actuelle les données démographiques et l'histoire des communes de France, base dite de « Cassini ». Il spécifie également le schéma de la base de données GeoPeople, et doit accueillir les données d'EHESS ainsi que les fonds géographiques créés par Hervé Le Bras, les fonds dits « Le Bras ». Enfin, il décrit les modalités de migration de Cassini vers GeoPeople.

Pour répondre à des besoins de généricté qui ont été défendus dans un précédent projet de EHESS, GéoNomenclature , mené par Jean-Pierre Pélissier et Claude Motte, la base de données doit être en mesure d'accueillir l'histoire de plusieurs nomenclatures territoriales (de différents types : administratives, financières, religieuses, etc.) qui peuvent avoir existé de façon concomitante. Nous proposons ici un schéma basé sur un paradigme identitaire, comme l'est la base Cassini actuelle, mais qui permet de prendre en compte les nouveaux besoins identifiés : génération d'un code unique par entité territoriale, et intégration de différentes nomenclatures territoriales.

## SOMMAIRE

<b>Intégration des données démographiques .....</b>	1
<b>Migration de la base de données Cassini vers GeoPeuple .....</b>	1
1.    Introduction : de la nécessité d'une migration .....	3
2.    Etude de Cassini.....	4
2.1.    Contenu de Cassini .....	4
2.1.    Le schéma de Cassini .....	8
2.1.1.    Partie entités .....	8
2.1.2.    Partie mutations.....	21
2.1.3.    Partie hiérarchie.....	26
3.    Le projet GeoNomenclature .....	31
3.1.    Rappel des besoins exprimés .....	31
3.2.    Spécification du code historique.....	32
3.3.    Conception de la base de données .....	36
3.3.1.    Première hypothèse .....	36
3.3.1.    Seconde hypothèse .....	39
3.3.2.    Gestion des nomenclatures .....	41
3.3.3.    La gestion du temps de validité.....	42
3.3.4.    Les métadonnées de provenance .....	45
3.3.1.    La gestion du versionnement de l'information (temps de transaction) .....	46
3.3.1.    Toponymes et chef-lieux .....	46
3.4.    Compatibilité avec d'autres bases existantes.....	48
4.    Implémentation et migration de la base de données.....	49
4.1.    Description des relations .....	49
4.1.1.    Les relations temporelles.....	49
4.1.1.    Les tables de métadonnées .....	50
4.1.2.    Les tables de nomenclature .....	52
4.1.3.    Les tables géographiques.....	60
4.2.    Migration de Cassini vers GeoPeuple.....	61
4.2.1.    Correspondance des types .....	61
4.2.2.    Analyse des toponymes .....	61
4.2.3.    Migration des événements.....	64
4.2.4.    Création des limites des communes .....	64
4.2.5.    Déterminer les communes des zones de saisie .....	68
4.2.1.    Procédure de migration .....	69
5.    Annexes.....	71
5.1.    Liste complète des zonages définis pour GeoNomenclature .....	71
5.2.    Analyse des cas de limites manquantes .....	72
5.2.1.    Limites actuelles non importées .....	72
5.2.2.    Erreurs détectées dans les géométries anciennes.....	73
6.    Bibliographie.....	77

# 1. Introduction : de la nécessité d'une migration

Pour les objectifs de GeoPeople, il est nécessaire de récupérer l'histoire des communes, les données démographiques qui sont attachées à ces communes, mais également de leur associer une empreinte spatiale, empreinte spatiale qui peut évoluer au cours du temps du fait des remembrements administratifs.

La base de données de EHESS, dite base « Cassini », contient l'ensemble des données relatives à l'histoire des communes, ainsi que les données démographiques. Cette base de données relationnelle, implémentée sous Access ne peut pas accueillir les limites spatiales, car il n'existe pas cette possibilité de gestion de la composante spatiale avec Access. Il est donc nécessaire de migrer la base de données Cassini dans une nouvelle base, implémentée sous Postgres ou Oracle, qui gèrent la dimension spatiale des données.

Mais par ailleurs, il s'avère que de nouveaux besoins ont été spécifiés en matière de gestion des nomenclatures territoriale dans un projet GeoNomenclature, et la migration des données est aussi l'opportunité de constituer un nouveau modèle répondant à ces besoins.

Dans un premier temps, nous rappelons les spécificités de la base Cassini, décrite en détail, afin que les subtilités de vocabulaire liées à l'histoire des communes soient bien saisies. Nous résumons dans un deuxième temps les nouveaux besoins qui ont été spécifiés, ou ceux qui sont ressortis de l'analyse de la base de données Cassini, puis proposons un modèle de données capable de répondre à ces besoins. Enfin, le plan détaillé de la migration et de la constitution de la base de données est documenté.

## 2. Etude de Cassini

L'objectif est de cette étude est de comprendre les données afin de définir dans la partie 4.2 de cette étude, page 61, les champs utiles pour renseigner le nouveau schéma.

L'étude de Cassini suppose d'avoir compris la nomenclature établie par EHESS pour analyser l'évolution des communes. Cette nomenclature est décrite dans plusieurs documents, et cette étude s'est appuyée sur le livre et le CD-ROM « COMMUNES d'HIER et d'AUJOURD'HUI, Dictionnaire d'histoire administrative », 2003, ed. INED, auteurs : Claude Motte, Isabelle Séguy, Christine Théré

L'étude de la base de données Cassini repose aussi sur la documentation fournie par Pascal Cristofoli :

- un schéma du modèle, « Cassini\_MCD.emf »,
- un document, inachevé, qui précise certaines lacunes de la base : « analyse\_BASE\_CASSINI.doc ».

La base, dont nous avons récupéré le schéma et les données, sert de support au site Web de EHESS : <http://cassini.ehess.fr/cassini/fr/html/index.htm>.

### 2.1. Contenu de Cassini

Une commune est définie par un territoire géographique, un nom et une population<sup>1</sup>, mais a aussi comme attributs un chef-lieu, un statut administratif (commune simple, chef-lieu de canton, sous-préfecture, ou préfecture) et une appartenance à une entité supérieure.

Les changements sur les communes sont enregistrés comme des événements, suivant la nomenclature suivante :

- **Création** à partir d'autres communes qui cèdent du terrain/des sections/des hameaux. Les communes qui cèdent du terrain subissent une *distraction*. Cette opération correspond à un morcellement du terrain et augmente le nombre de communes.
- **Absorption** : une commune est toujours absorbée ou absorbante. Est considérée comme absorbante la commune qui possède ou récupère le chef-lieu. Cette opération diminue le nombre de communes.
- **Transfert de chef-lieu**
- **Changement de nom**
- **Rétablissement** : cette opération augmente le nombre de communes

Les cinq types d'événements peuvent avoir lieu de façon simple ou simultanée. Comme on postule qu'il n'y a *jamais de fusion* avec dilution de l'identité (Cf. livre p44), il est possible d'avoir un *rétablissement* (l'opération inverse) : on peut faire revivre des communes ou entités ayant disparues (c.a.d. ayant été absorbées antérieurement).

L'analyse des historiens qui est dans la BD Cassini est très fine. On peut se reporter par exemple à la Figure 1 qui illustre l'ensemble des cas considérés dans les notices des communes.

---

<sup>1</sup> Ceci ne signifie pas que le recensement identifie une commune.

## Rappel des légendes, des abréviations et de la typographie

### Codification spatiale

- « A » signifie : une commune Absorbante
- « a » signifie : une commune absorbée
- « C » signifie : une commune Créeée
- « c » signifie : une commune cédante
- « A2 » signifie : absorbe deux communes
- « a2 » signifie : absorbée avec une autre commune
- « C3 » signifie : créée avec deux autres communes à partir d'une même commune
- « c3 » signifie : cède des territoires pour former trois nouvelles communes
- « Ap » signifie : absorbe, avec d'autres, une commune
- « ap » signifie : absorbée simultanément par plusieurs communes
- « Cp » signifie : créée à partir de parcelles appartenant à plusieurs communes
- « cp » signifie : cède, avec d'autres communes, un territoire pour former une nouvelle commune

### Codification temporelle

- « A.A » signifie : absorbe deux communes à des dates différentes
- « A.a » signifie : absorbe une commune, puis est absorbée à une date ultérieure
- « A/c » signifie : absorbe une commune et cède une portion de son territoire pour former une nouvelle commune

### Transfert d'un ancien territoire communal

- « d » signifie : cède, à une autre, le territoire d'une commune qu'elle avait précédemment absorbée
- « R » signifie : reçoit le territoire d'une commune auparavant absorbée par une autre
- « t » signifie : absorbée par une commune puis transférée dans une autre

### Transfert de chef-lieu communal

- « tcl » signale : la commune qui perd le chef-lieu
- « Ch-l » signale : l'ancienne commune qui reçoit le chef-lieu

### Rappel de la signification de la typographie utilisée pour l'index des toponymes

- le gras signale : une commune actuelle
- le maigre signale : une ancienne commune
- le maigre-italique signale : une variante du nom d'une commune (actuelle ou ancienne)

**Figure 1. Liste des abréviations utilisées dans les notices communales. Extrait CD-ROM "Communes d'hier et d'aujourd'hui".**

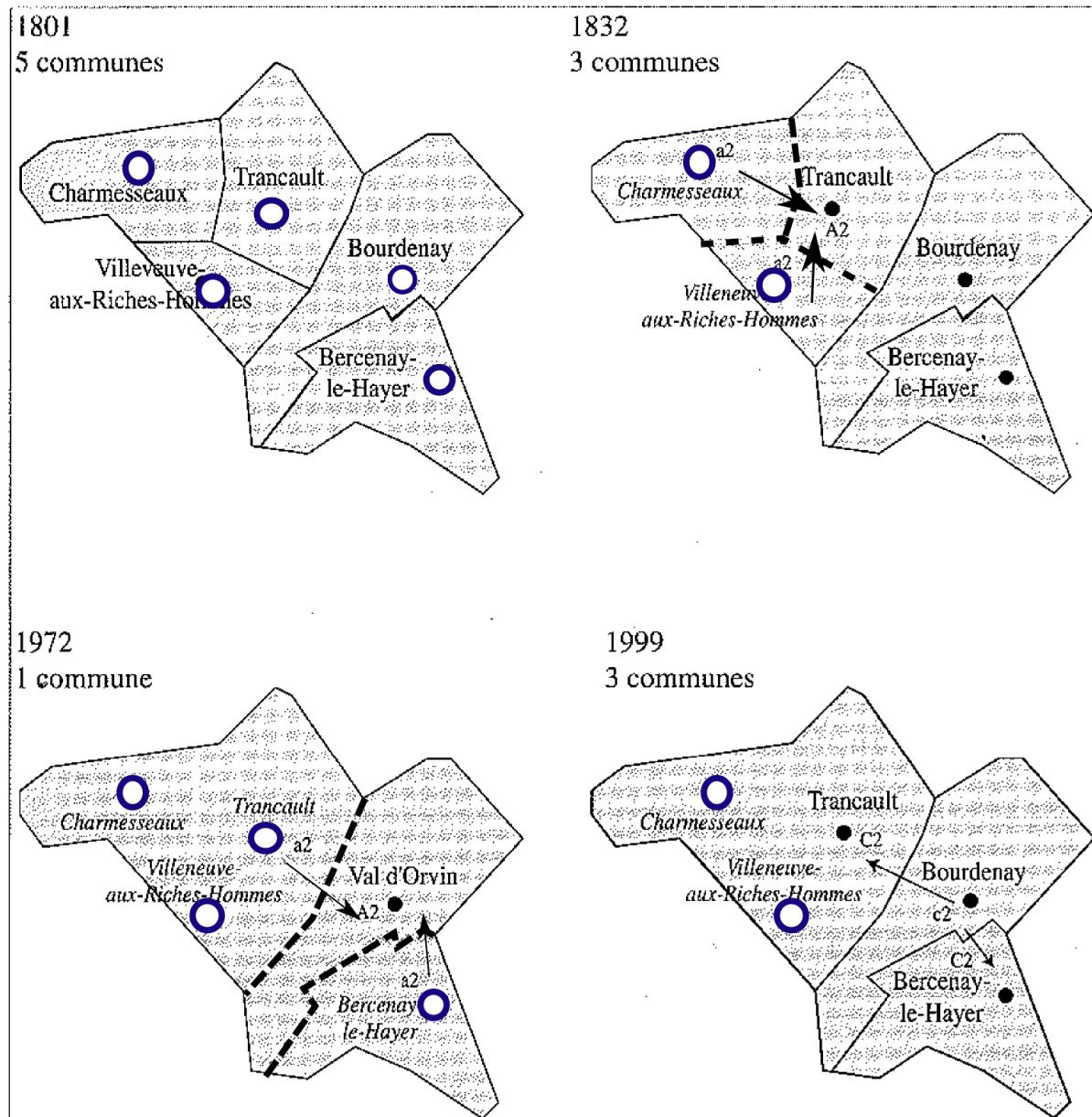
Voici un exemple de notice publiée dans le livre « communes d'hier et d'aujourd'hui » p54 et p55.

- Charmesseaux
- Trancault
- Villeneuve-Aux-Riches-Hommes
- Val d'Orvin
- Bourdenay
- Bercenay-le-Hayer

Bercenay-le-Hayer (10.2.13.038)	sup. : "?" ha	alt. : 99/238 m	1806 : 283 hab.	1999 : ... hab.	a2.C2
événements : a.1972, (avec Trancault) Bourdenay ; C.1999, Bourdenay					
<i>nom (1801)</i> : Bercenay-le-Hayer					
<i>dép. (1801)</i> : Aube					
<i>arr. (1801)</i> : Nogent-sur-Seine					
<i>cant. (1801)</i> : Marcilly-le-Hayer					
Bourdenay (10.2.13.054)	sup. : 6.012 ha	alt. : 94/186 m	1806 : 244 hab.	1999 : 402 hab.	A2.c2
événements : A.1972, Bercenay-le-Hayer/Trancault ; c.1999, Bercenay-le-Hayer/Trancault					
<i>nom (1801)</i> : Bourdenay ; 1972, Val-d'Orvin ; 1999, Bourdenay					
<i>dép. (1801)</i> : Aube					
<i>arr. (1801)</i> : Nogent-sur-Seine					
<i>cant. (1801)</i> : Marcilly-le-Hayer					
Charmesseaux (10.054)	sup. : ... ha	alt. : ... / ... m	1806 : 25 hab.	1999 : ... hab.	a2
événements : a.1832, (avec Villeneuve-aux-Riches-Hommes) Trancault					
<i>nom (1801)</i> : Charmesseaux					
<i>dép. (1801)</i> : Aube					
<i>arr. (1801)</i> : Nogent-sur-Seine					
<i>cant. (1801)</i> : Marcilly-le-Hayer					
Trancault (10.2.13.383)	sup. : "? ha	alt. : 86/166 m	1806 : 239 hab.	1999 : ... hab.	A2.a2.C2
événements : A.1832, Charmesseaux/Villeneuve-aux-Riches-Hommes ; a.1972, (avec Bercenay-le-Hayer) Bourdenay ; C.1999, (avec Trancault) Bourdenay					
<i>nom (1801)</i> : Trancault					
<i>dép. (1801)</i> : Aube					
<i>arr. (1801)</i> : Nogent-sur-Seine					
<i>cant. (1801)</i> : Marcilly-le-Hayer					
Villeneuve-aux-Riches-Hommes (10.054)	sup. : ... ha	alt. : ... / ... m	1806 : 22 hab.	1999 : ... hab.	a2
événements : a.1832, (avec Charmesseaux) Trancault					
<i>nom (1801)</i> : Villeveuve-aux-Riches-Hommes					
<i>dép. (1801)</i> : Aube					
<i>arr. (1801)</i> : Nogent-sur-Seine					
<i>cant. (1801)</i> : Marcilly-le-Hayer					

Figure 2. Extrait du livre, p54 : notices des communes citées.

La Figure 3 illustre ces notices et montre l'évolution de ces communes.



**Figure 3. Extrait du livre, p55 : illustration de l'évolution des communes citées.**

Les nomenclatures considérées dans la base de Cassini sont :

- La hiérarchies des départements-arrondissements<sup>2</sup>-cantons-communes ayant existé de 1790 à 1801
- La hiérarchies des départements-arrondissements-cantons-communes ayant existé de 1801 à de nos jours. Les régions actuelles sont mentionnées, mais leur histoire n'est pas enregistrée.

On remarque que, dans le cas des communes, ces deux hiérarchies sont traitées comme si elles étaient identiques car un identifiant unique (*num\_ordre*) permet de repérer la même commune au cours de ces deux périodes. En revanche, pour les entités de niveau supérieur (cantons, arrondissements, départements), elles sont dupliquées (voire multipliées) avec une première période d'existence allant de 1793 à 1801 puis une seconde partant de 1801 jusqu'au prochain remembrement affectant ou non l'entité<sup>3</sup>.

<sup>2</sup> Jusqu'à 1795, les arrondissements sont aussi appelés *districts*.

<sup>3</sup> Pour signaler la permanence de validité d'une entité (existante à l'heure actuelle), l'année 2500 est utilisée.

## 2.1. Le schéma de Cassini

Le schéma de Cassini est documenté dans le document « Cassini\_MCD.emf », dont la Figure 4 donne une vue générale.

Globalement, le schéma se décompose en trois parties :

- La partie centrale qui décrit les entités territoriales de niveau communes, avec leur toponymes, leur existence, et les recensements ;
- La partie qui décrit la hiérarchie des entités territoriales (les subdivisions, et les relations entre les entités) ;
- La partie qui décrit les événements (ou mutations) qui interviennent sur les communes.

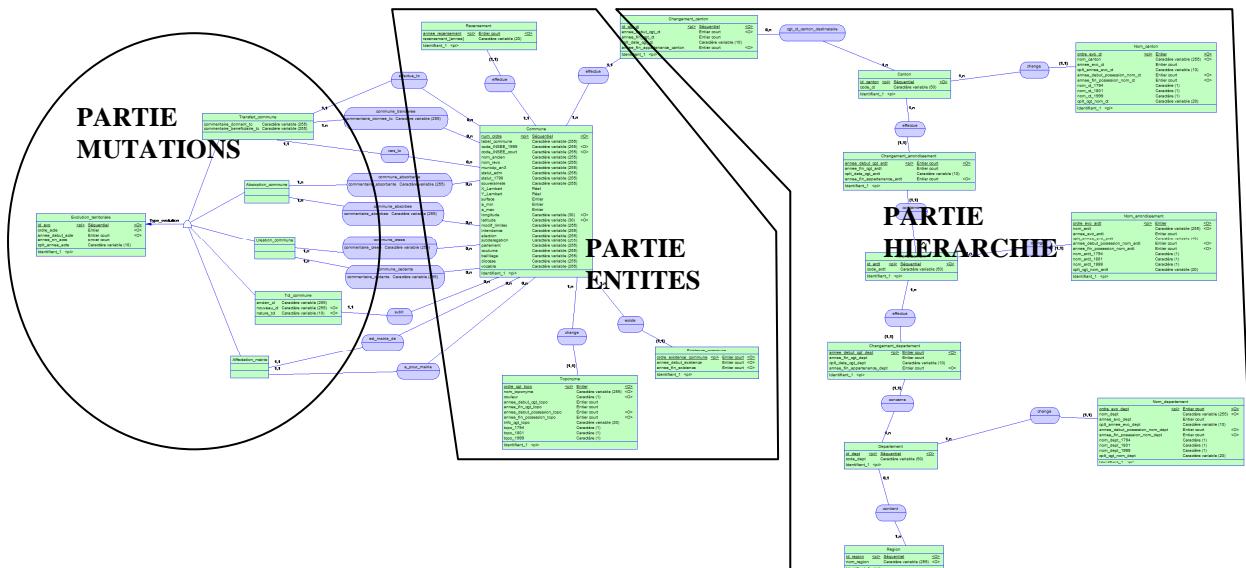


Figure 4. Vue générale du modèle de Cassini.

La lecture du modèle ne permet pas de comprendre immédiatement le sens de certains champs. Il s'avère de plus que certaines tables de la base de données diffèrent du modèle. Afin de faciliter la compréhension de la migration, nous documentons donc de nouveau ce schéma.

Quelques remarques valables pour tout le modèle :

- Tous les champs « sel » de type TINYINT(1) fonctionnent comme des booléens : une valeur -1 semble indiquer FAUX, et VRAI sinon. Ces champs servent simplement à faciliter certaines interrogations sous ACCESS : ils sont inutiles.
- Un certains nombre de champs sont indexés (certainement pour accélérer les recherches) avec le mot clé KEY. <http://dev.mysql.com/doc/refman/5.0/fr/create-table.html>
- En revanche, aucune clé étrangère (FOREIGN KEY) n'a été spécifiée.

### 2.1.1. Partie entités

#### 2.1.1.1. Table COMMUNE

Dans “analyse\_BASE\_CASSINI.doc”, le numéro d’ordre (NUM\_ORDRE) est documenté comme « identifiant des communes repérée par Claude Motte dans la base cassini : (commune ayant eu une

existence entre 1790 et 2006) ». C'est donc l'équivalent d'un *serial id*. Lorsque le num\_ordre est supérieur à 60000, l'entité est une portion infra-communale d'une commune actuelle, systématiquement. C'est généralement une ancienne commune ou communauté de communes entre 1790 et 1801.

Pour les communes ayant été absorbées et n'existant plus aujourd'hui, leur « code\_insee\_court », comme leur code «code\_insee\_1999 » vaut le « code\_insee\_court » de leur commune de rattachement existant à l'heure actuelle. Pour les communes existantes à l'heure actuelle, leur code\_insee\_1999 est de longueur supérieur à 7 digits.

champ	type	contrainte	signification
num_ordre	serial	NOT NULL, PK	Id : numéro identifiant cette commune ou portion communale de 1790 à de nos jours. 10140
nom_1801	varchar(255)	default NULL	Valeur de « nom_toponyme » qui apparaît avec la valeur topo_1801 à 'E' dans la table TOPOONYME Conflans
label_commune	varchar(255)	default NULL	Valeur du « nom_toponyme » le plus récent. Albertville
code_insee_1999	varchar(255)	default NULL	Ce code correspond au code de l'INSEE pour les communes qui ont été recensées par l'INSEE ('73 1 98 011'), ou bien à leur code court si les communes n'existent plus ('73 011' pour Saint Sigismond rattaché à Albertville).
code_insee_court	varchar(50)	default NULL	Code insee court de la commune '73 011'
sel	tinyint(1)	NOT NULL	Toutes les entrées de la table ont ce champ valué à '-1' : inutile
nom_ancien	varchar(255)	default NULL	Seulement 3 entrées <sup>4</sup> ont un nom_ancien non null. Il correspond à un nom antérieur à la révolution française (1793) <sup>5</sup>
nom_revo	varchar(255)	default NULL	Nom donné à la commune à la révolution française (1793). Ce champ répète une entrée de la table TOPOONYME. 'Roc-Libre'
municip_an3	varchar(255)	default NULL	Indique dans quel recensement ses habitants sont comptés en 1793. Par exemple '1793, Conflent' Ou "1793, Chazey, Rothond et Bons" : en

<sup>4</sup> select num\_ordre, nom\_1801, code\_insee\_1999, nom\_ancien from COMMUNE where nom\_ancien is not null

num_ordre	nom_1801	code_insee_1999	nom_ancien
18372	Laire	25 2 30 322	Tavay (Expilly), Tavey (Cassini)
32501	Saint-Jean-de-Niort	01 2 19 361	Gourdans
198	Aibre	25 2 30 008	Eubre (Expilly)

<sup>5</sup> Ce champ servira donc à renseigner une subdivision antérieure à 1790.

			1793, les habitants de Chazy, Rothond et Bons sont comptés ensemble.
nom_an3	varchar(255)	default NULL	Nom qui apparait pour le recensement 1793. 'Conflent' Ou "Rothonod"
statut_adm	varchar(255)	default NULL	'chef-lieu de canton', 'préfecture', 'Sous-préfecture', ou 'commune' ou 'ancienne commune', ou <b>d'autres valeurs, moins standards</b> , mais qui donnent des informations intéressantes à retraiter pour l'histoire des chef-lieux d'entités supérieures. La liste énumérés de tous les statuts présents est donnée par la requête suivante : select statut_adm, count(*) from commune group by statut_adm order by count DESC
statut_1790	varchar(255)	default NULL	Statut en 1790 'chef-lieu de canton (anII)' ou autre ... <b>Très souvent vide.</b> La liste énumérés de tous les statuts présents est donnée par la requête suivante : select statut_1790, count(*) from commune group by statut_1790 order by count DESC
souverainete	varchar(255)	default NULL	Etat et département d'appartenance entre 1790 et 1800. '1789, royaume de France ; 1790, Seine et Oise' "1789, Etats Sardes (duché de Savoie) ; 1792, Mont Blanc"
x_lambert	double(15,5)	default NULL	Centre, projeté en coordonnées Lambert 915908 <b>0 pour valeur qui n'a pas lieu d'être</b>
y_lambert	double(15,5)	default NULL	Centre, projeté en coordonnées Lambert 2083200 <b>0 pour valeur qui n'a pas lieu d'être</b>
surface	integer	default NULL	Surface en hectares (ha) 1754
a_min	integer	default NULL	Altitude minimum de la commune (m) 328
a_max	integer	default NULL	Altitude maximum de la commune(m) 2030
longitude	varchar(30)	default NULL	Longitude du centre " 6°23m33s E" <b>... pour valeur qui n'a pas lieu d'être.</b>
latitude	varchar(30)	default NULL	Latitude du centre " 45°40m33s N" <b>... pour valeur qui n'a pas lieu d'être.</b>

modif_limites	varchar(255)	default NULL	Prévu pour décrire une modification des limites impliquant un échange de parcelle x, où x n'aurait pas été une ancienne commune. TOUJOURS NULL.
region_1999	varchar(255)	default NULL	Région supérieure en 1999 "Rhône-Alpes" <b>'...' pour valeur qui n'a pas lieu d'être.</b>
dept_1999	varchar(255)	default NULL	Département supérieur en 1999 "Savoie" <b>'...' pour valeur qui n'a pas lieu d'être.</b>
ardt_1999	varchar(255)	default NULL	Arrondissement supérieur en 1999 "Albertville" <b>'...' pour valeur qui n'a pas lieu d'être.</b>
ct_1999	varchar(255)	default NULL	Canton supérieur en 1999 "Albertville-N., -S." <b>'...' pour valeur qui n'a pas lieu d'être.</b>
c_suite	longtext		Texte abrégé <sup>6</sup> qui décrit l'événement qui a modifié la commune. Par exemple : "A.1835, L'Hôpital ; A.1964, Saint-Sigismond" Ce champ est extrait du fichier Excel ayant servi à constituer la base Cassini.
c_suite_1790	longtext		Texte abrégé qui décrit l'événement qui a modifié la commune entre 1790 et 1801. Par exemple : « a.<1790-1794>, Romescamps » Ce champ est extrait du fichier Excel ayant servi à constituer la base Cassini.
c_suite_dev	longtext		Texte qui décrit l'événement qui a modifié la commune. Par exemple : "Absorbe en 1835, L'Hôpital ; Absorbe en 1964, Saint-Sigismond" Ce champ sert à vérifier les informations décrites par la partie mutations du modèle.
c_suite_1790_dev	longtext		Texte qui décrit l'événement qui a modifié la commune entre 1790 et 1801. Par exemple : « réunie entre 1790-1794, à Romescamps» Ce champ sert à vérifier les informations décrites par la partie mutations du modèle.
chf_lieu_communal	varchar(255)	default NULL	<b>Très souvent vides (seulement 849 enregistrements sur un total de 43878),</b> ces champs renseignent sur l'appartenance à des subdivisions autres qu'administratives. Les communes renseignées semblent faire partie de
intendance	varchar(255)	default NULL	
election	varchar(255)	default NULL	
subdelegation	varchar(255)	default NULL	
parlement	varchar(255)	default NULL	
coutume	varchar(255)	default NULL	

<sup>6</sup> Les abréviations correspondent à la codification des mutations listée dans la Figure 1.

bailliage	varchar(255)	default NULL	l'intendance de « Moulins » dans l'Allier, le plus souvent.
diocese	varchar(255)	default NULL	
vocable	varchar(255)	default NULL	

Tableau 1. La table COMMUNE.

#### 2.1.1.2. *Table RECENSEMENT et GESTION\_RECENSEMENT*

Les valeurs des recensements (nombre d'habitants) des communes sont toutes enregistrées dans la table RECENSEMENT. Chaque enregistrement pour une commune contient la valeur de ses recensements pour les années allant de 1793 à 1999. Le champ « recensement\_autre » correspond en fait au recensement de l'année 2006.

Si les valeurs renseignées ne sont pas des entiers (*integer*) pour les colonnes RECENSEMENT\_XXX, c'est en raison de l'utilisation de la codification décrite dans le Tableau 2 pour expliquer les valeurs manquantes. Mais des valeurs NULL existent quand même.

Pour les départements de Nice et ceux appartenant à l'Italie, ainsi que pour l'Alsace-Lorraine, les valeurs sont manquantes ou correspondent à des dates différentes durant certaines périodes parce que ces territoires ne faisaient plus partie de la France à ces périodes.

Code	Signification
0 hab.	commune inhabitée
...	commune n'existant pas à cette date
adm.	commune recensée avec une autre
lac.	commune oubliée sur la publication du recensement
ill.	information illisible
vide	information non disponible actuellement
abs.	document (ou pages) disparu(es)

Tableau 2. Codification expliquant les valeurs de recensement absentes.

champ	type	contrainte	signification	Italie, Nice	Alsace-Lorraine
num_ordre	integer	Default NULL, PK	Numero d'ordre de la commune recensée.		
RECENSEMENT_1793	varchar(20)	default NULL	Valeur en 1793		
RECENSEMENT_1800	varchar(20)	default NULL	Valeur en 1800		
RECENSEMENT_1806	varchar(20)	default NULL	Valeur en 1806		
RECENSEMENT_1821	varchar(20)	default NULL	Valeur en 1821	1822	
RECENSEMENT_1831	varchar(20)	default NULL	Valeur en 1831	Abs.	
RECENSEMENT_1836	varchar(20)	default NULL	Valeur en 1836	1838	
RECENSEMENT_1841	varchar(20)	default NULL	Valeur en 1841	Abs.	
RECENSEMENT_1846	varchar(20)	default NULL	Valeur en 1846	1848	
RECENSEMENT_1851	varchar(20)	default NULL	Valeur en 1851	Abs.	
RECENSEMENT_1856	varchar(20)	default NULL	Valeur en 1856	1858	
RECENSEMENT_1861	varchar(20)	default NULL	Valeur en 1861		
RECENSEMENT_1866	varchar(20)	default NULL	Valeur en 1866		
RECENSEMENT_1872	varchar(20)	default NULL	Valeur en 1872		1870
RECENSEMENT_1876	varchar(20)	default NULL	Valeur en 1876		1875
RECENSEMENT_1881	varchar(20)	default NULL	Valeur en 1881		1880
RECENSEMENT_1886	varchar(20)	default NULL	Valeur en 1886		1885

RECENSEMENT_1891	varchar(20)	default NULL	Valeur en 1891		1890
RECENSEMENT_1896	varchar(20)	default NULL	Valeur en 1896		1895
RECENSEMENT_1901	varchar(20)	default NULL	Valeur en 1901		1900
RECENSEMENT_1906	varchar(20)	default NULL	Valeur en 1906		1905
RECENSEMENT_1911	varchar(20)	default NULL	Valeur en 1911		1910
RECENSEMENT_1921	varchar(20)	default NULL	Valeur en 1921		
RECENSEMENT_1926	varchar(20)	default NULL	Valeur en 1926		
RECENSEMENT_1931	varchar(20)	default NULL	Valeur en 1931		
RECENSEMENT_1936	varchar(20)	default NULL	Valeut en 1936		
RECENSEMENT_1946	varchar(20)	default NULL	Valeur en 1946		
RECENSEMENT_1954	varchar(20)	default NULL	Valeur en 1954		
RECENSEMENT_1962	varchar(20)	default NULL	Valeur en 1962		
RECENSEMENT_1968	varchar(20)	default NULL	Valeur en 1968		
RECENSEMENT_1975	varchar(20)	default NULL	Valeur en 1975		
RECENSEMENT_1982	varchar(20)	default NULL	Valeur en 1982		
RECENSEMENT_1990	varchar(20)	default NULL	Valeur en 1990		
RECENSEMENT_1999	varchar(20)	default NULL	Valeur en 1999		
RECENSEMENT_Autre	varchar(20)	default NULL	<b>Valeur de 2006</b>		

**Tableau 3. La table RECENSEMENT.**

La table GESTION\_RECENSEMENT n'est pas documentée. Elle est par ailleurs exemptes d'informations intéressantes, en dehors de la date de recensement, qui peut aisément être déduite du nom des champs de la table RECENSEMENT. Nous ne nous étendons donc pas sur sa définition.

#### **2.1.1.3. Table TOPOONYME ET NOM\_XXX**

La table TOPOONYME indique le nom d'une commune. Le changement de nom est daté par une période et non pas un instant : une année de début et une année de fin de changement sont spécifiées ; ces deux dates permettent de signaler une incertitude sur une date, et de décrire qu'un changement a eu lieu entre deux dates. Deux années de changement de toponyme sont précisées : en effet, on peut avoir une imprécision sur la période de changement (entre la date 'ANNEE\_DEBUT\_CGT\_TOPO' et la date 'ANNEE\_FIN\_CGT\_TOPO'). Si par contre la date est connue de façon certaine, seule le champ « ANNEE\_DEBUT\_CGT\_TOPO » est rempli.

Une information sous la forme d'un commentaire accompagne la modification du toponyme et renseigne sur la source du toponyme : 'ACTE' signifie que le toponyme a été trouvé dans un acte, il est donc officiel.

La période de validité du toponyme est également renseignée indiquant les dates de début et fin de possession de ce toponyme. Toutes les valeurs de « ANNEE\_DEBUT\_CGT\_TOPO » et « annee\_DEBUT\_POSSESSION\_TOPO » sont égales. Lorsque le toponyme est encore d'actualité, « ANNEE\_FIN\_POSSESSION\_TOPO » vaut 2500. « ANNEE\_FIN\_POSSESSION\_TOPO » vaut aussi 2500 si la commune en question a disparu. Dans certains cas (détaillés ci-dessous), ce n'est pas la date renseignée dans « ANNEE\_DEBUT\_CGT\_TOPO » qui sert de date de début de validité.

Les tables NOM\_CANTON, NOM\_ARRONDISSEMENT et NOM\_DEPARTEMENT ressemblent fortement à la table TOPOONYME et contiennent la presque même information, mais pour ces autres entités respectivement. La seule différence est que pour les noms des communes, un code couleur est associé à chaque toponyme.

- en rouge ('R') : nom d'une commune actuelle
- en noir ('N') : dernier nom connu d'une commune disparue

- en vert ('V') : dénominations intermédiaires
- Un index existe sur le champ « nom\_toponyme ».

champ	type	contrainte	signification
num_ordre	integer	FK sur la table COMMUNE, Default NULL	Id de la commune référencée qui change de nom
ORDRE_CGT_TOPO	integer	default NULL	Numero identifiant ce changement de nom de façon unique pour la commune référencée
NOM_TOPONYME	varchar(255)	default NULL	Nouveau nom
COULEUR	varchar(1)	default NULL	'N', 'V', ou 'R'
ANNEE_DEBUT_CGT_TOPO	smallint(5)	default NULL	Change entre la date x (ou à la date x précisément).
ANNEE_FIN_CGT_TOPO	smallint(5)	default NULL	Et la date y. Y vaut null si on est sur que le changement a eu lieu la date x
ANNEE_DEBUT_POSSESSION_TOPO	smallint(5)	default NULL	Début date de validité du nom
ANNEE_FIN_POSSESSION_TOPO	smallint(5)	default NULL	Fin de validité du nom
INFO_CGT_TOPO	varchar(20)	default NULL	Commentaire : renseigne sur la source (ACTE) où est apparu ce toponyme.
Sel	tinyint(1)	NOT NULL	Inutile (juste pour access)
TOPO_1794	varchar(1)	default NULL	Est nom en 1794 (anII)
TOPO_1801	varchar(1)	default NULL	'E' : Est nom en 1801 (BO) 'C' : est cité avec d'autres communes 'N' : n'existe pas au BO, mais présente au recensement Voir p194 de « Communes d'hier et d'aujourd'hui »
TOPO_1999	varchar(1)	default NULL	Est nom en 1999 (INSEE_99)

Tableau 4. La table TOPOONYME.

Les champs « TOPO\_1794 », « TOPO\_1801 », et « TOPO\_1999 » constituent une vérification par rapport à 3 sources, **officielles** : an II, le BO et la base de données INSEE\_99. Un caractère ('E') indique si ce toponyme était celui de la commune en 1794, 1801 et 1999. Ce sont des noms officiels. Les noms officiels sont aussi ceux dont « INFO\_CGT\_TOPO » vaut « ACTE ».

Ces informations servent à déterminer la date de début de validité d'un toponyme.

Déterminer la date de début de validité :

- si TOPO\_1794 = 'E', alors c'est 1793
- si TOPO\_1801 = 'E', alors c'est 1801
- si TOPO\_1999 = 'E', on n'a pas toujours la date de changement de nom.
- si INFO\_CGT\_TOPO = 'ACTE', alors c'est ANNEE\_DEBUT\_POSSESSION\_TOPO

Ainsi, pour "Lou Mas de Labastide" et "L'Abadie", le champ « annee\_debut\_possession\_topo » vaut 1793, mais le site Web indique que "L'Abadie" est le toponyme valide à partir de 1801 et "Lou Mas de Labastide" à partir de 1793.

Il semble (en vérifiant avec les sources INSEE 2010) que certains noms sont officiels et devraient être notés 'E' dans topo\_1999. C'est le cas d'Aboën par exemple, (Figure 5).

**Le nom révolutionnaire** se lit dans le champ « nom\_revo » de la table COMMUNE. Une **variante** Est une variante un toponyme qui n'est ni révolutionnaire, ni officiel, et qui joue double emploi à côté d'un nom officiel. Par exemple, Aas est une variante de Aast, entre 1793 et 1861, comme le confirme le site Web (Figure 6).

<b>Aboën</b>	
Le nom	
ancien(s)	information non disponible actuellement
nom(s) rév.	information non disponible actuellement
an II	1793, Aboin et Salineaux
Bulletin des Lois	
	<ul style="list-style-type: none"> <li>• 1801, ...</li> <li>• 1872, Aböen</li> <li>• Aboën</li> </ul>

Figure 5. Extrait du site Web : Aboën.

<b>Eaux-Bonnes</b>	
Le nom	
ancien(s)	information non disponible actuellement
nom(s) rév.	information non disponible actuellement
an II	1793, Aas
Bulletin des Lois	
	<ul style="list-style-type: none"> <li>• 1801, Aast</li> <li>• Aas</li> <li>• 1861, Eaux-Bonnes</li> </ul>

Figure 6. Extrait du site Web : Eaux-Bonnes.

Cependant, l'exemple de Vensat laisse perplexe car il semble être une exception aux règles précédemment énoncées, voir Figure 7 :

num_ordre	ordre_cgt_topo	nom_toponyme	couleur	annee_debut_cgt_topo	annee_fincgt_topo	annee_debut_possession_topo	annee_fin_possession_topo	info_cgtopo	topo_1794	topo_1801	topo_1999
39661	130	Vensat	R	1800		1800	2500	ACTE			E
39661	120	Saint-Jean-de-Vensat	V			1793	1800				
39661	110	Veussat	V			1793	1800			E	
39661	100	Saint Jean de Venssat	V	1793		1793	1800	ACTE	E		

Tableau 5. Extrait de la base Cassini pour la commune Vensat.

D'après ces règles :

- « Saint Jean de Venssat » serait le nom officiel en 1793 (an II)
- « Veussat » devrait être un toponyme révolutionnaire.

- « Vensat » serait est le nom officiel depuis 1800 à de nos jours.
- « Saint-Jean-de-Vensat » serait une variante existant entre 1793 et 1800

Or le site Web ne mentionne pas « Veussat » ni « Saint-Jean-de-Vensat » (voir la Figure 7).

<b>Vensat</b>	
Le nom	
ancien(s)	information non disponible actuellement
nom(s) rév.	information non disponible actuellement
an II	1793, Saint Jean de Venssat
	1800, Vensat
Bulletin des Lois	
	1801, ...

**Figure 7. Extrait du site Web : Vensat.**

champ	type	contrainte	signification
ID_CANTON	integer	FK sur la table CANTON, Default NULL	Id du canton référencé qui change de nom
ORDRE_EVO_CT	integer	default NULL	Numero identifiant ce changement de nom de facon unique pour le canton référencé
NOM_CANTON	varchar(255)	default NULL	Nouveau nom
ANNEE_EVO_CT	smallint(5)	default NULL	Année du changement de nom. Souvent NULL.
CPLT_ANNEE_EVO_CT	varchar(10)	default NULL	Renseigné avec « av » pour avant (imprécision sur la date), sinon NULL.
ANNEE_DEBUT_POSSESSION_NOM_CT	smallint(5)	default NULL	Début date de validité du nom. Non NULL
ANNEE_FIN_POSSESSION_NOM_CT	smallint(5)	default NULL	Fin de validité du nom
CPLT_CGT_NOM_CT	varchar(20)	default NULL	Vaut 'Tcl' quand il est présent
Sel	tinyint(1)	NOT NULL	Inutile (juste pour access)
NOM_CT_1794	varchar(1)	default NULL	Toujours NULL
NOM_CT_1801	varchar(1)	default NULL	Toujours NULL
NOM_CT_1999	varchar(1)	default NULL	Toujours NULL

**Tableau 6. La table NOM\_CANTON.**

Certains champs de la table TOPOONYME semblent équivalents à ceux de NOM\_CANTON :

ANNEE\_DEBUT\_CGT\_TOPO ≈ ANNEE\_EVO\_CT

ANNEE\_DEBUT\_POSSESSION\_TOPO ≈ ANNEE\_DEBUT\_POSSESSION\_NOM\_CT

ANNEE\_FIN\_POSSESSION\_TOPO ≈ ANNEE\_FIN\_POSSESSION\_NOM\_CT

Sel ≈ Sel

TOPO\_1794 ≈ NOM\_CT\_1794

TOPO\_1801 ≈ NOM\_CT\_1801

TOPO\_1999 ≈ NOM\_CT\_1999

INFO\_CGT\_TOPO ≈ CPLT\_CGT\_NOM\_CT

Un index existe sur le champ « nom\_canton ».

Au lieu d'utiliser deux colonnes pour signifier l'imprécision sur une date de changement, la date est spécifiée dans le champ « ANNEE\_EVO\_CT » et le champ « CPLT\_ANNEE\_EVO\_CT » permet de préciser si c'est avant ou après, semble-t-il. On ne trouve que deux entrées non nulles pour ce champ (Tableau 7).

De plus « ANNEE\_EVO\_CT » est toujours égale à « ANNEE\_DEBUT\_POSSESSION\_NOM\_CT », lorsqu'elle est renseignée, et vaut nulle sinon.

<b>id_canton</b>	<b>ordre_evo_ct</b>	<b>nom_canton</b>	<b>annee_evo_ct</b>	<b>cplt_annee_evo_ct</b>	<b>annee_debut_possession_nom_ct</b>	<b>annee_fin_possession_nom_ct</b>
5036	130	Salies	1943	av	1943	1958
7013	120	La Salvetat	1943	av	1943	1958

**Tableau 7. Entrées utilisant le champ "CPLT\_ANNEE\_EVO\_CT".**

Le champ « cplt\_cgt\_nom\_ct » signale si le changement de nom du canton correspond à un transfert de chef-lieu. Dans ce cas, on trouve presque toujours une commune du même nom qui est chef-lieu du canton à cette date. Sauf trois exceptions (Tableau 8).

<b>id_canton</b>	<b>nom_canton</b>
7988	Mézières (Ch.-l.: Charleville-Mézières)
9553	Cherbourg-Octeville-S.O.
10373	Baraqueville-Sauveterre

**Tableau 8. Cantons avec Tcl mais pas d'entrée correspondante dans NOM\_TOPONYME.**

champ	type	contrainte	signification
ID_ARDT	integer	FK sur la table CANTON, Default NULL	Id de l'arrondissement référencé qui change de nom
ORDRE_EVO_ARDT	integer	default NULL	Numero identifiant ce changement de nom de façon unique pour l'arrondissement référencé
NOM_ARDT	varchar(255)	default NULL	Nouveau nom
ANNEE_EVO_ARDT	smallint(5)	default NULL	Début date de validité du nom (peut être NULL)
CPLT_ANNEE_EVO_ARDT	varchar(10)	default NULL	Précise la date de changement avec le mois
ANNEE_DEBUT_POSSESSION_NOM_ARDT	smallint(5)	default NULL	Début date de validité du nom, non NULL.
ANNEE_FIN_POSSESSION_NOM_ARDT	smallint(5)	default NULL	Fin de validité du nom
CPLT_CGT_NOM_ARDT	varchar(20)	default NULL	Vaut 'Tcl' quand il est présent
Sel	tinyint(1)	NOT NULL	Inutile (juste pour access)
NOM_ARDT_1794	varchar(1)	default NULL	Toujours NULL
NOM_ARDT_1801	varchar(1)	default NULL	Toujours NULL
NOM_ARDT_1999	varchar(1)	default NULL	Toujours NULL

**Tableau 9. La table NOM\_ARRONDISSEMENT.**

Certains champs de la table TOPOONYME semblent équivalents à ceux de NOM\_ARRONDISSEMENT :

ANNEE\_DEBUT\_CGT\_TOPO ≈ ANNEE\_EVO\_ARDT

ANNEE\_DEBUT\_POSSESSION\_TOPO ≈ ANNEE\_DEBUT\_POSSESSION\_NOM\_ARDT

ANNEE\_FIN\_POSSESSION\_TOPO ≈ ANNEE\_FIN\_POSSESSION\_NOM\_ARDT

Sel ≈ Sel

TOPO\_1794 ≈ NOM\_ARDT\_1794

TOPO\_1801 ≈ NOM\_ARDT\_1801

TOPO\_1999 ≈ NOM\_ARDT\_1999

INFO\_CGT\_TOPO ≈ CPLT\_CGT\_NOM\_ARDT

Un index existe sur le champ « nom\_arondissement ».

Le champ « cplt\_cgt\_nom\_ardt » signale si le nom du arrondissement correspond à un transfert de chef-lieu. Dans ce cas, on trouve toujours une commune du même nom qui est chef-lieu de l'arrondissement à cette date (elle est donc sous-préfecture<sup>7</sup>). Sans exception.

Au lieu d'utiliser deux colonnes pour signifier l'imprécision sur une date de changement, la date est spécifiée dans le champ « ANNEE\_EVO\_ARDT » et le champ « CPLT\_ANNEE\_EVO\_ARDT » permet de

<sup>7</sup> Elle peut aussi être préfecture si elle est chef-lieu du département.

préciser la date avec le mois. On ne trouve que deux entrées non nulles pour ce champ (Tableau 7). De plus « ANNEE\_EVO\_ARDT » est toujours égale à « ANNEE\_DEBUT\_POSSESSION\_NOM\_ARDT », lorsqu'elle est renseignée, et vaut nulle sinon.

<b>id_ar dt</b>	<b>ordre_evo_a rdt</b>	<b>nom_arde</b>	<b>annee _evo_ar dt</b>	<b>cplt_ann ee _evo_ar dt</b>	<b>annee_debut_ possession_nom_ ardt</b>	<b>annee_fin_ possession_nom_ ardt</b>
857	130	Châlons-sur-Marne	1997	avr	1997	1997
857	140	Châlons-en-Champagne	1997	déc	1997	2500

Tableau 10. Entrées utilisant le champ "CPLT\_ANNEE\_EVO\_ARDT".

champ	type	contrainte	signification
ID_DEPT	integer	FK sur la table CANTON, Default NULL	Id du département référencé qui change de nom
ORDRE_EVO_DEPT	integer	default NULL	Numéro identifiant ce changement de nom de façon unique pour le département référencé
NOM_DEPT	varchar(255)	default NULL	Nouveau nom
ANNEE_EVO_DEPT	smallint(5)	default NULL	Début date de validité du nom
CPLT_ANNEE_EVO_DEPT	varchar(10)	default NULL	Toujours NULL
ANNEE_DEBUT_POSSESSION_NOM_DEPT	smallint(5)	default NULL	Début date de validité du nom, non NULL.
ANNEE_FIN_POSSESSION_NOM_DEPT	smallint(5)	default NULL	Fin de validité du nom
CPLT_CGT_NOM_DEPT	varchar(20)	default NULL	Toujours NULL
Sel	tinyint(1)	NOT NULL	Inutile (juste pour access)
France	tinyint(1)	NOT NULL	Signale si un département faisait partie de la France à l'époque mentionnée (peut expliquer l'absence de recensement ou une date de recensement différente).
NOM_DEPT_1794	varchar(1)	default NULL	Toujours NULL
NOM_DEPT_1801	varchar(1)	default NULL	Toujours NULL
NOM_DEPT_1999	varchar(1)	default NULL	Toujours NULL

Tableau 11. La table NOM\_DEPARTEMENT.

Certains champs de la table TOPO NYME semblent équivalents à ceux de NOM\_DEPARTEMENT:

ANNEE\_DEBUT\_CGT\_TOPO ≈ ANNEE\_EVO\_DEPT

ANNEE\_DEBUT\_POSSESSION\_TOPO ≈ ANNEE\_DEBUT\_POSSESSION\_NOM\_DEPT

ANNEE\_FIN\_POSSESSION\_TOPO ≈ ANNEE\_FIN\_POSSESSION\_NOM\_DEPT

Sel ≈ Sel

TOPO\_1794 ≈ NOM\_DEPT\_1794

TOPO\_1801 ≈ NOM\_DEPT\_1801

TOPO\_1999 ≈ NOM\_DEPT\_1999

INFO\_CGT\_TOPO ≈ CPLT\_CGT\_NOM\_DEPT

Un index existe sur le champ « nom\_departement ».

Le champ « France » est un champ supplémentaire par rapport aux autres tables, qui signale si le département en question faisait partie de la France durant la période mentionnée : à -1, oui, à 0, non. Elle permet de suivre les évolutions des communes et cantons sortis de France qui la réintégreront plus tard.

id_dept	ordre_evo_dept	nom_dept	annee_evo_dept	annee_debut_possession_nom_dept	annee_fin_possession_nom_dept	france
26	110	Moselle (Allemagne)	1871	1871	1919	0
30	110	Haut-Rhin (Allemagne)	1871	1871	1919	0
51	100	Haute-Savoie (Etats Sardes)	1816	1816	1860	0
64	110	Bas-Rhin (Allemagne)	1871	1871	1919	0
104	110	Alpes-Maritimes (Etats Sardes)	1814	1814	1860	0
109	100	Savoie (Etats Sardes)	1816	1816	1860	0
142	100	Alpes-Maritimes (Italie)	1860	1860	1947	0
		Alpes-Maritimes (principauté de Monaco)				
145	100	Alpes-Maritimes (principauté de Monaco)	1814	1814	1848	0
147	100	Moselle (Forêts-Luxembourg)	1795	1795	1801	0
148	100	Alpes-Maritimes (ville libre sous protection du roi de Sardaigne)	1848	1848	1861	0

Tableau 12. Liste des départements qui ne faisaient pas partie de la France aux époques mentionnées.

« ANNEE\_EVO\_ARDT » est toujours égale à « ANNEE\_DEBUT\_POSSESSION\_NOM\_ARDT », et non nulle.

#### 2.1.1.4. Table EXISTENCE\_COMMUNE

Cette table donne la période de validité d'une entité. Si « annee\_fin\_existence » vaut 2500, c'est que l'entité est toujours valide. Sinon, il faut rechercher son num\_ordre dans la table EVOLUTION\_TERRITORIALE pour connaître la raison de sa disparition.

champ	type	contrainte	signification
num_ordre	integer	FK sur la table COMMUNE, NOT NULL	Id de la commune référencée qui change de nom
ORDRE_EXISTENCE_COMMUNE	smallint(5)	NOT NULL	Numero identifiant cette période d'existence de façon unique pour la commune référencée
ANNEE_DEBUT_EXISTENCE	smallint(5)	default NULL	Début de cette période d'existence
ANNEE_FIN_EXISTENCE	smallint(5)	default NULL	Fin de cette période d'existence

**Tableau 13. La table EXISTENCE\_COMMUNE.**

Cette table contient des erreurs. Par exemple, la commune Lardenne (id : 62427) a disparu par absorption en 1794 avec Toulouse, mais a pour annee\_fin\_existence la valeur 2500. Ceci peut se vérifier avec les requêtes suivantes :

```
select * from commune where num_ordre=62427
select * from EXISTENCE_COMMUNE e where e.num_ordre=62427
```

### 2.1.2. Partie mutations

La partie mutations décrit l'ensemble des mutations qui peuvent intervenir sur une commune et modifier son existence. Toutes les mutations sont des évolutions territoriales (qu'elles spécialisent) et à ce titre, elles réfèrent toutes la table EVOLUTION\_TERRITORIALE.

#### 2.1.2.1. Table EVOLUTION\_TERRITORIALE

La principale table est la table EVOLUTION\_TERRITORIALE qui décrit un événement (une mutation) liant plusieurs entités territoriales. Elle est référencée par toutes les tables de cette partie car elle renseigne sur les dates de l'évolution.

champ	type	contrainte	signification
Id_evo	Integer (serial)	PK, UNIQUE, NOT NULL	Id de l'événement
ORDRE_ACTE	integer	default NULL	Numero identifiant cette période d'existence de façon unique pour la commune référencée
ANNEE_DEBUT_ACTE	smallint(5)	default NULL	Début de l'événement
ANNEE_FIN_ACTE	smallint(5)	default NULL	Fin de l'événement
CPLT_ANNEE_ACTE	Varchar(10)	default NULL	Prend la valeur 'av', 'av.' ou NULL le plus souvent. Lorsque le champ est renseigné, ANNEE_FIN_ACTE vaut toujours nul.

**Tableau 14. La table EVOLUTION\_TERRITORIALE.**

La date de l'événement peut être imprécise : si les deux dates sont renseignées (debut et fin),

l'événement a eu lieu entre telle et telle date. Les dates des actes concordent avec les dates d'existence de la table EXISTENCE\_COMMUNE :

- Pour une création, la commune créée a pour début d'existence une date comprise entre la date de début et de fin de l'acte. Si la fin de l'acte n'est pas renseignée, c'est la date de début de l'acte qui marque le début de l'existence.
- Pour une absorbtion, la commune absorbée a pour fin d'existence une date comprise entre le début et la fin de l'acte. Si la fin de l'acte n'est pas renseignée, c'est la date de début de l'acte qui marque la fin de l'existence.

Requête pour lister l'ensemble des absorbtions avec les dates (d'acte, d'existence) des entités absorbées :

```
select distinct c.label_commune, a.num_ordre_absorbee, a.id_evo, ct.num_ordre_absorbante,
e.annee_debut_existence, e.annee_fin_existence, et.annee_debut_acte, et.annee_fin_acte,
c.c_suite_1790_dev, c.c_suite_dev
from commune_absorbante ct, commune_absorbee a, EXISTENCE_COMMUNE e, EVOLUTION_TERRITORIALE
et, commune c
where ct.id_evo=a.id_evo and ct.id_evo=et.id_evo and e.num_ordre=a.num_ordre_absorbee
and ((e.annee_fin_existence = et.annee_debut_acte and et.annee_fin_acte is null) or
(e.annee_fin_existence >= et.annee_debut_acte and e.annee_fin_existence <= et.annee_fin_acte
and et.annee_fin_acte is not null))
and c.num_ordre=a.num_ordre_absorbee order by id_evo
```

Requête pour lister l'ensemble des créations avec les dates (d'acte, d'existence) des entités créées :

```
select c.label_commune, cnew.num_ordre_creee, cnew.id_evo, cold.num_ordre_cedante,
et.annee_debut_acte, et.annee_fin_acte, e.annee_debut_existence, e.annee_fin_existence,
c.c_suite_1790_dev, c.c_suite_dev
from commune_cedante cold, commune_creee cnew, EVOLUTION_TERRITORIALE et, EXISTENCE_COMMUNE e,
COMMUNE c
where cold.id_evo=cnew.id_evo and cold.id_evo=et.id_evo and e.num_ordre=cnew.num_ordre_creee
and ((e.annee_debut_existence=et.annee_debut_acte and et.annee_fin_acte is null)
or (e.annee_debut_existence>=et.annee_debut_acte and et.annee_fin_acte is not null))
and c.num_ordre=cnew.num_ordre_creee order by id_evo
```

### 2.1.2.2. Table TRANSFERT\_COMMUNE

La table TRANSFERT\_COMMUNE mentionne les événements ayant conduit une commune à céder une section de son territoire à une autre (mais sans que cela ait donné lieu ni à la création d'une commune ni à la suppression d'une commune).

En revanche, le nom de la commune donneuse ou prenante peut avoir changé si le chef-lieu était sur la section transférée, ou si cette modification a été l'occasion de changer le nom de l'une des communes impliquées. Cet événement modifie forcément les limites du territoire des communes. Ne sont mentionnés que les modifications modifiant le recensement communal.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_DONNEUSE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui cède du territoire
NUM_ORDRE_BENEFICIAIRE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui récupère du territoire
COMMENTAIRE_BENEFICIAIRE_TC	Varchar(255)	Default NULL	Commentaire à afficher pour la commune bénéficiaire. <b>Toujours NULL.</b>
COMMENTAIRE_DONNANT_TC	Varchar(255)	Default NULL	Commentaire à afficher pour la commune donneuse. <b>Toujours NULL sauf pour deux entrées.</b>

Tableau 15. La table TRANSFERT\_COMMUNE.

La table COMMUNE\_TRANSFEREE mentionne les événements ayant conduit une commune à être transférée entièrement vers une autre. Dans un même événement de transfert, une ou plusieurs communes peuvent être transférées entre deux communes. Le transfert ne détermine pas forcément la date de disparition de l'entité transférée. Par exemple, Etoquet, disparue par absorption en 1794 dans Livry a été ensuite transférée de Livry à Caumont-l'Eventé. Ceci distingue le transfert de l'absorption : le transfert a lieu sur des communes qui sont déjà disparues (car elles sont section d'autres communes), alors que l'absorption correspond à leur disparition « initiale ».

champ	type	contrainte	signification
Id_evo	Integer	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_TRANSFEREE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui est transférée
COMMENTAIRE_DONNEE_TC	Varchar(255)	Default NULL	Commentaire à afficher pour la commune transférée. <b>Vaut toujours NULL</b>

Tableau 16. La table COMMUNE\_TRANSFEREE.

#### 2.1.2.3. Tables ABSORPTION\_COMMUNE, COMMUNE\_ABSORBANTE, COMMUNE\_ABORBEE

La table ABSORPTION\_COMMUNE mentionne les événements ayant conduit une commune à disparaître par absorption dans une ou plusieurs communes. Cet événement provoque la suppression d'une commune. Lors d'un événement, une commune peut absorber plusieurs communes à la fois. Cet événement modifie forcément les limites du territoire de la commune qui absorbe l'autre en s'agrandissant.

Exemple : La Garrigue-Haute (id : 50015) est absorbée par trois communes ([Campouriez](#) / [Montézic](#) / [Saint-Amans-des-Cots](#)) en 1833.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution

Tableau 17. La table ABSORPTION\_COMMUNE.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_ABSORBANTE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui absorbe

COMMENTAIRE_ABSORBANTE	Varchar(255)	Default NULL	Commentaire à afficher pour la commune absorbante. Vaut souvent NULL et l'information n'est pas exploitée sur le site Web.
------------------------	--------------	--------------	--

Tableau 18. La table COMMUNE\_ABSORBANTE.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_ABSORBEE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui est absorbée
COMMENTAIRE_ABSORBEE	Varchar(255)	Default NULL	Commentaire à afficher pour la commune absorbée. Vaut souvent NULL et l'information n'est pas exploitée sur le site Web.

Tableau 19. La table COMMUNE\_ABSORBEE.

#### 2.1.2.4. Table CREATION\_COMMUNE, COMMUNE\_CEDANTE, COMMUNE\_CREEE

La table CREATION\_COMMUNE mentionne les événements ayant conduit à la création d'une commune grâce à la cession de parcelles d'autres communes qui n'ont pas forcément disparues. Par exemple Abancourt (id : 9) est créée en 1823 à partir de Blargies (id : 29536) et Romescamps (id : 4466), et Blargies continue d'exister aujourd'hui.

Cet événement provoque la création d'une commune et modifie forcément les limites du territoire des communes qui cèdent du territoire (tout ou partie) pour la création de la commune. Les communes cédantes continuent d'exister, sauf, cas rare (Tableau 20), lorsqu'elles sont dans la foulée absorbée ou réunie à une autre commune.

label_commune	c_suite_dev	num_ordre_cedante	annee_debut_acte	annee_debut_existence	annee_fin_existence
Ceignac	Absorbe en 1829, Calmont / Magrin / Milhac / Naves ; rÉunie en 1832, À Calmont // donne À Calmont les h. Magrin / Milhac // donne À Manzac le h. Naves	9728	1832	1793	1832
Saint-	Absorbe en 1847, ChÀteaubernard ; cÈde en 1867,	3325	1867	1793	1867

Martin	Châteaubernard / réunie à Cognac	9			
Bonnipaire	cède en 1848, Combrimont ; réunie en 1848, à Bertrimoutier	4885	<b>1848</b>	1793	<b>1848</b>
Alciette-Bascassan	cède en 1842, (avec Ahaxe-Alciette-Bascassan / Aincille / Bustince-Iriberry / Caro / Lecumberry / Mendive / Saint-Jean-le-Vieux / Saint-Michel) Estrenéuby ; réunie en 1842, à Ahaxe-Alciette-Bascassan	393	<b>1842</b>	1793	<b>1842</b>
Esperons	cède en 1861, (avec Saint-Loubouer) Eugénie-les-Bains ; réunie en 1861, à Eugénie-les-Bains	13020	<b>1861</b>	1793	<b>1861</b>
Bézacoul	cède en 1833, (avec Paulinet) Teillet ; réunie en 1833, à Teillet	4139	1833	1793	1833
Ampiac	réunie en 1837, (avec Abbas / Agnac / Ayssials-Lagarrigue / Castan / ls-Bonnecombe) à Druelle	12253	1837	1793	1837

Tableau 20. Liste des communes disparues en cédant du territoire.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution

Tableau 21. La table CREATION\_COMMUNE.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_CEDANTE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui cède du territoire, unique dans cet événement
COMMENTAIRE_CEDANTE	Varchar(255)	Default NULL	Commentaire à afficher pour la commune cédante. Vaut souvent NULL et l'information n'est pas exploitée sur le site Web.

Tableau 22. La table COMMUNE\_CEDANTE.

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_CREEE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui est créée, unique dans cet événement
COMMENTAIRE_CREEE	Varchar(255)	Default NULL	Commentaire à afficher pour la commune créée. Vaut souvent NULL et l'information n'est pas exploitée sur le site Web.

Tableau 23. La table COMMUNE\_CREEE.

#### 2.1.2.5. Table TCL\_COMMUNE

La table TCL\_COMMUNE décrit les transferts de chef-lieux à l'intérieur d'une commune. Le chef-lieu d'une commune change suite à la montée en puissance d'un de ses hameaux. Dans ce cas il y a transfert de chef-lieu vers le hameau dominant, et le hameau est soit une ancienne commune soit un

lieu-dit (non connu comme entité territoriale). Dans ce dernier cas, il faut considérer le chef-lieu comme une entité de niveau infra-communal.

Voir p197,198, 199 de « Communes d'hier et d'aujourd'hui ».

champ	type	contrainte	signification
Id_evo	Integer (serial)	Defaut NULL, PK, UNIQUE	Id de l'événement d'évolution
NUM_ORDRE_TRANSFERANTE	integer	FK sur COMMUNE, default NULL	identifiant de la commune qui perd son chef-lieu (si tcl) ou qui reçoit le chef- lieu (si Ch-l).
ANCIEN_CL	varchar(255)	default NULL	Nom de l'ancien chef-lieu (pas forcément renseigné).
NOUVEAU_CL	Varchar(255)	Default NULL	Nom du nouveau chef-lieu (toujours renseigné)
NATURE_TCL	Varchar(10)	Default NULL	Description du transfert : 'tcl' ou 'Ch-l' (jamais NULL) 'tcl' : la commune perd le chef-lieu 'Ch-l' : la commune reçoit le chef-lieu (elle est rétablie).

Tableau 24. La table TCL\_COMMUNE.

Pour les communes, le chef-lieu est un lieu géographique qui prend la forme :

- d'un lieu connu par sa latitude et sa longitude,
- et/ou d'une entité connue (ancienne commune devenue hameau),
- et/ou d'une entité dont on ne connaît que le nom (par exemple, 'les Junies' dans le cas de Almont-les-Junies). C'est pourquoi les chef-lieux sont désignés par des noms (chaîne de caractères) dans la base, et non des entités (des identifiants d'entités).

La table TCL\_COMMUNE permet de renseigner sur les deux derniers cas.

#### 2.1.2.6. Table AFFECTATION\_MAIRIE

La table affectation\_mairie est vide. Elle aurait pu être utilisée à la place de TCL\_COMMUNE. L'analyse des exemples nous fait penser qu'elle est inutile.

#### 2.1.3. Partie hiérarchie

La partie hiérarchie décrit la structure administrative du territoire : une commune appartient à un canton qui appartient à un arrondissement qui appartient à un département, qui lui-même fait partie d'une région. Les changements d'appartenances entre entités sont retracés dans le modèle, excepté pour les liens départements-régions, car l'histoire des régions n'est pas retracée.

On observe que les tables canton, et arrondissement ont un champ LABEL\_XXX qui renseigne sur leur nom, mais ce nom ne concorde pas toujours avec les tables NOM\_XXX. Par exemple, pour les cantons, 1820 cantons n'ont pas le même toponyme que leur champ LABEL\_CANTON. De même, 144 arrondissements n'ont pas pour nom leur label.

### 2.1.3.1. Table CANTON

La table CANTON recense l'ensemble des cantons observés entre 1790 et 2006 : 11428 au total. Le document « analyse\_BASE\_CASSINI.doc » signale que dès qu'un remembrement des cantons était effectué, on considère que ce sont des nouveaux cantons. Le code INSEE renseigné n'est pas celui des cantons (reste à saisir) mais celui de la commune chef-lieu du canton.

champ	type	contrainte	signification
Id_CANTON	Int(10), (serial)	NOT NULL, PK, UNIQUE	Id du canton
CODE_CT	Varchar(50)	default NULL	Code du canton
EXISTE	Tinyint(1)	NOT NULL	si 0, alors le canton n'existe plus en 1999.
num_dept	varchar(50)	default NULL	Numéro du département d'appartenance du canton.
code_INSEE	varchar(50))	default NULL	Code INSEE du chef-lieu du canton. Pour l'instant, celui de la commune.
LABEL_CANTON	Varchar(255)	Default NULL	Nom du canton (pas forcément concordant avec son toponyme)
Sel	Tinyint(1)	NOT NULL	inutile

Tableau 25. La table CANTON.

champ	type	contrainte	signification
ID_CGT_CT	Int(10), (serial)	Defaut NULL, PK, UNIQUE	Id du changement d'appartenance
ID_CANTON	Int(10)	default NULL	Id du canton

Tableau 26. La table CGT\_CT\_CANTON\_DESTINATAIRE.

champ	type	contrainte	signification
ID_CGT_CT	Int(10), (serial)	NOT NULL, PK, UNIQUE	Id du changement d'appartenance
NUM_ORDRE	Int(10)	default NULL	Id de la commune
ANNEE_DEBUT_CGT_CT	smallint(5)	default NULL	Année de début d'appartenance au canton. Borne inférieure de la période de changement si année_fin_cgt_ct n'est pas nul.
ANNEE_FIN_CGT_CT	smallint(5)	default NULL	Année de début d'appartenance : borne supérieure. souvent NULL
CPLT_DATE_CGT_CT	varchar(10)	default NULL	Précise la date de début d'appartenance. NULL ou 'av' pour « avant »
ANNEE_FIN_APPARTENANCE_CANTON	smallint(5)	default NULL	Année de fin d'appartenance au canton. 2500 si appartenance toujours valide.

Tableau 27. La table CHANGEMENT\_CANTON.

La table CHANGEMENT\_CANTON lie toute commune à son canton : elle appartient au canton ‘x’ entre à la date « annee\_debut\_cgt\_ct » et la date « annee\_fin\_appartenance ». La date « annee\_debut\_cgt » peut être précisée :

- soit comme une date imprécise, avec une date de fin de changement « annee\_fin\_cgt\_ct » qui limite la période de changement,
- soit avec un commentaire dans « cplt\_date\_cgt\_ct » qui indique si le changement a lieu avant la date mentionnée.

Le canton ‘x’ n’est pas référencé directement : il faut faire une jointure sur la table CGT\_CT\_CANTON\_DESTINATAIRE et CANTON pour trouver le canton d’appartenance durant la période [≈annee\_debut\_cgt\_ct, annee\_fin\_appartenance].

En effet, en général, une commune n’appartient qu’à un seul canton. Cependant, dans le cas des cantons urbains, comme Caen-1 à Caen-6 qui divisent la ville de Caen en 6 parties depuis 1973, la commune de Caen est enregistrée 6 fois comme appartenant à chacun des cantons urbains de Caen. Le rôle de la table CGT\_CT\_CANTON\_DESTINATAIRE est d’enregistrer ces relations n-m entre cantons et communes.

Enfin, la table TABLE\_CHEF\_LIEU\_CANTON est vide. Elle est inutile.

#### **2.1.3.2. Table ARRONDISSEMENT**

La table ARRONDISSEMENT recense l’ensemble des arrondissements observés entre 1790 et 2006. Le document « analyse\_BASE\_CASSINI.doc » signale que dès qu’un remembrement des arrondissements était effectué, on considère que ce sont des nouveaux arrondissement. Le code INSEE renseigné n’est pas celui des arrondissements (reste à saisir) mais celui de la commune chef-lieu de l’arrondissement.

champ	type	contrainte	signification
Id_ARDT	Int(10), (serial)	NOT NULL, PK, UNIQUE	Id de l’arrondissement
CODE_ARDT	Varchar(50)	default NULL	Code de l’arrondissement
EXISTE	Tinyint(1)	NOT NULL	si 0, alors l’arrondissement n’existe plus en 1999.
num_dept	varchar(50)	default NULL	Numéro du département d’appartenance de l’arrondissement.
code_INSEE	varchar(50))	default NULL	Code INSEE de la commune chef-lieu de l’arrondissement (sous-préfecture)
LABEL_ARDT	Varchar(255)	Default NULL	Nom de l’arrondissement
Sel	Tinyint(1)	NOT NULL	inutile

**Tableau 28. La table ARRONDISSEMENT.**

La table CHANGEMENT\_ARRONDISSEMENT définit les changements d’arrondissement pour un canton. Un canton ne peut appartenir qu’à un seul arrondissement la même année à cause de la contrainte d’unicité sur le triplet (id\_canton, id\_arde, annee\_debut\_cgt\_arde).

champ	type	contrainte	signification
ID_CANTON	integer	default NULL	Id du canton
ID_ARDT	integer	default NULL	Id de l’arrondissement du canton
ANNEE_DEBUT_CGT_ARDT	smallint(5)	default NULL	Année de début d’appartenance à

			l'arrondissement. Borne inférieure de la période de changement si annee_fin_cgt_arde n'est pas nul.
ANNEE_FIN_CGT_ARDT	smallint(5)	default NULL	Année de début d'appartenance : borne supérieure. souvent NULL
CPLT_DATE_CGT_ARDT	varchar(10)	default NULL	Précise la date de début d'appartenance. NULL ou 'av' pour « avant »
ANNEE_FIN_APPARTENANCE_ARDT	smallint(5)	default NULL	Année de fin d'appartenance à l'arrondissement. 2500 si appartenance toujours valide.

**Tableau 29. La table CHANGEMENT\_ARRONDISSEMENT.**

La table SOUS\_PREFECTURE (équivalent de chef-lieu d'arrondissement) est vide. La table est inutile.

#### 2.1.3.3. *Table DEPARTEMENT*

La table DEPARTEMENT recense l'ensemble des départements observés entre 1790 et 2006. Le document « analyse\_BASE\_CASSINI.doc » signale qu'il existe des départements qui ont disparus ou qui correspondent à des territoires étrangers à l'heure actuelle.

champ	type	contrainte	signification
Id_DEPT	Int(10), (serial)	NOT NULL, PK, UNIQUE	Id du département
ID_REGION	integer	default NULL	Id de la région d'appartenance du département.
CODE_DEPT	Varchar(50)	default NULL	Code du département
EXISTE	Tinyint(1)	NOT NULL	si 0, alors le département n'existe plus en 1999.
Sel	Tinyint(1)	NOT NULL	inutile

**Tableau 30. La table DEPARTEMENT.**

La table CHANGEMENT\_DEPARTEMENT définit les changements de département pour un arrondissement. Un arrondissement ne peut appartenir qu'à un seul département la même année à cause de la contrainte d'unicité sur le triplet (id\_arde, id\_dept, annee\_debut\_cgt\_dept).

champ	type	contrainte	signification
ID_ARDT	integer	default NULL	Id de l'arrondissement
ID_DEPT	integer	default NULL	Id du département de l'arrondissement
ANNEE_DEBUT_CGT_DEPT	smallint(5)	default NULL	Année de début d'appartenance au département. Borne inférieure de la période de changement si

			annee_fin_cgt_dept n'est pas nul.
ANNEE_FIN_CGT_DEPT	smallint(5)	default NULL	Année de début d'appartenance : borne supérieure. souvent NULL
CPLT_DATE_CGT_DEPT	varchar(10)	default NULL	Précise la date de début d'appartenance. NULL ou 'av' pour « avant »
ANNEE_FIN_APPARTENANCE_DEPT	smallint(5)	default NULL	Année de fin d'appartenance au département. 2500 si appartenance toujours valide.

**Tableau 31. La table CHANGEMENT\_DEPARTEMENT.**

La table PREFECTURE (équivalent de chef-lieu de département) est vide. La table est inutile.

#### **2.1.3.4. Table REGION**

La table région liste les régions, simplement pour des raisons de regroupement statistique. Leur histoire n'a pas été traitées contrairement aux autres entités territoriales

champ	type	contrainte	signification
ID_REGION	integer (serial)	Defaut NULL, PK, UNIQUE	Id de la region
NOM_REGION	varchar(255)	default NULL	Nom de la region

**Tableau 32. La table REGION.**

### 3. Le projet GeoNomenclature

#### 3.1. Rappel des besoins exprimés

Cette partie rappelle quels sont les besoins qui avaient été définis par le projet GéoNomenclature, détaillés dans le rapport « Géonoménclature historique des lieux habités », Décembre 2003, de Jean-Pierre Pélissieret Claude Motte, disponible sur Internet (<http://www.archivesdefrance.culture.gouv.fr/static/1079>). Il s'agit en résumé de concevoir une structure de données **extrêmement pérenne** capable d'accueillir toutes les *nomenclatures* territoriales ayant existé, existantes ou qui existeront sur le territoire français.

Les *nomenclatures* (ou *division*) se présentent sous la forme d'un ensemble de découpages territoriaux organisés de façon hiérarchique (ou quasi-hiéarchique, ou pas du tout). Chaque nomenclature est non seulement un outil d'analyse scientifique mais aussi un outil d'action de gouvernance [Guibert 1971], que celle-ci soit du domaine du religieux, de l'administratif ou de l'économique : il existe ainsi des nomenclatures religieuses, administratives, fiscales, etc, que l'on peut encore spécialiser sous le terme de *subdivision* : la subdivision administrative française 1801 (celle de l'INSEE) ou la subdivision Nomenclature des Unités Territoriales Statistiques (NUTS) définie par l'Europe.

Chaque nomenclature définit différents *niveaux*, ou partitionnements, de l'espace géographique étudié, appelées *maillage* ou *zonage* selon que ces partitions sont complètes ou bien incomplètes (dans ce second cas, la partition n'a pas obligation à couvrir complètement l'espace). Les *mailles* ou *zones* ou *entités territoriales* qui composent chaque zonage sont à la fois des parcelles de territoire et des territoires à part entière qui entretiennent entre eux des relations hiérarchiques (c'est-à-dire de gouvernance) mais également de compétition, qui font évoluer ces relations hiérarchiques.

Par ailleurs, les niveaux ne sont pas nécessairement strictement emboîtés, ni ne forment une partition complète et totale de l'espace. Lorsque, par exemple, un historien s'intéresse à la nomenclature française des unités fiscales à l'époque Moderne, selon les dictionnaires de Saugrain (1709 et 1720), il ne liste pas moins de 23 subdivisions de l'espace de niveau 2 (bailliages, comtés, diocèses civils, duchés, élections, etc.) qui cohabitent ensemble, et 38 subdivisions de l'espace de niveau 3, (dont les paroisses, précurseurs des communes actuelles [Chareille et al. 2004]). Il observe que, d'une part, ces subdivisions ne forment pas des partitions complètes et couvrantes de l'espace (l'ensemble des duchés ne couvre pas la France) et que, d'autre part, les entités entretiennent des relations d'appartenance multiples avec les unités de niveau supérieur.

Ces *entités* sont considérées comme fictives au départ, au sens que lui confère Barry Smith, (ce sont des "*flat objects*"), car leurs délimitations, et leurs attributs (noms, centre, statut) sont le produit d'une volonté humaine. Elles sont porteuses d'éléments statistiques associés à des documents signifiants pour l'histoire économique, sociale et politique des territoires sous-jacents, comme par exemple les registres de naissance, mariage et décès attachés aux communes. Au fil du temps, elles ne cessent d'être redéfinies. Et surtout, comme pour les communes, elles peuvent s'incarner plus profondément dans la réalité, et prendre « chair » sur le territoire, devenir des entités qui ont une identité propre, et auxquelles leurs habitants s'identifient.

Pour les historiens, comme les géographes, l'étude du territoire français qui s'inscrit dans de multiples subdivisions, administratives, religieuses, fiscales nécessite une identification précise des entités étudiées. Il s'agit par exemple de déterminer à qui une entité territoriale reverse ses impôts

par exemple, et/ou si elle connaît plusieurs tutelles simultanément. Grâce au suivi de ces changements d'appartenance, l'historien peut analyser des causes et des effets de conflits, politiques et/ou économiques. Tout ceci n'est possible que s'il existe **une codification claire et unique de cette entité**. En effet, il existe un certain nombre d'homonymie au niveau des entités territoriales de base (par exemple trois communes sont dénommées "Fontaine" en France en 2012), ainsi que dans l'intitulé des subdivisions auxquelles appartiennent ces entités : ainsi « paroisse » fait référence aussi bien à la paroisse religieuse de la période moderne qu'à celle de la période contemporaine (telle que définie par la [Commission sociale des évêques de France, 1998]) mais aussi à la paroisse administrative de la division financière de l'Ancien Régime. Egalement, lorsqu'une source délivre des informations sur la collecte ou le montant d'une imposition dans le cadre d'un « bailliage », il est nécessaire de déterminer s'il est question de l'entité « bailliage » relevant de la division fiscale ou du « bailliage » de la division judiciaire. Etablir une codification univoque des entités territoriales, être en mesure de reconnaître à quelle emprise spatiale, à quel nom, quel code, et quel statut il est fait référence est donc l'ambition du projet Géonoménclature.

Ce projet constitue une extension (temporelle, en amont de 1943, et dans les différentes divisions du territoire, fiscale, religieuse, judiciaire, militaire, géographique) du **Code Officiel Géographique** (COG) de l'INSEE (Institut National des Statistiques et Etudes Economiques). Le COG, document de référence qui rassemble les codifications (numérotations et libellés) des communes, des cantons, des arrondissements, des départements, des régions, des collectivités d'outre-mer, des pays et territoires étrangers [Lang, 2010]. L'INSEE diffuse également un historique des recompositions des communes sur son site : « *le fichier historique des communes relate les changements de nom, les créations, les disparitions et les changements de département intervenus pour la France métropolitaine et les départements d'outre-mer (DOM) depuis 1943 de façon exhaustive. Les autres modifications (échanges de parcelles, transferts de chef-lieu, changements d'appartenance à un échelon supracommunal ou créations de cantons infracommunaux) sont répertoriées depuis 1994. Par ailleurs, les échanges de parcelles sans incidence démographique pour lesquels la superficie des communes est inchangée ne sont pas enregistrés de manière exhaustive.* »<sup>8</sup>

Nous considérons ainsi que la commune est une entité territoriale, comme le canton, l'arrondissement, le département, et la région, que la nomenclature définie par l'INSEE est une nomenclature administrative parmi d'autres. Il s'agit de construire un système permettant d'enregistrer ces entités et leur évolutions, ainsi que de leur associer **un code historique unique**, représentatif de leur emprise territoriale, dénomination, appartenance, et statut.

### 3.2. Spécification du code historique

Le code historique doit être unique pour chaque version d'entité territoriale. Il est construit à partir d'un système similaire au code de l'INSEE : l'entité est numérotée de façon unique par rapport une unité de référence à laquelle est appartient. Comme les communes sont numérotés de façon unique par rapport à leur département. Par exemple, 38 130 est le code insee court de la commune d'Echirolles : 130<sup>ème</sup> commune du département de l'Isère (n°38).

Cependant, à la différence du code INSEE, le code géonoménclature intègre les éléments suivants : il indique également de quelle nomenclature il est question, et il précise le zonage de l'entité désignée. Ce code rend aussi compte des modifications survenues sur l'entité au cours du temps, depuis sa création (création dans la base s'entend), en précisant à la fois la nature du changement, et ensuite le numéro d'occurrence du changement de cette entité (premier changement, deuxième changement, troisième, etc.).

---

<sup>8</sup> Extrait le 8 mai 2012 de : [http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc\\_fhistor.htm](http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc_fhistor.htm)

La spécification du code géonomenclature est la suivante :

**Code1-code2#code3-code4#code5-code6-code7**  
ann-nn#nnn-nnnn#nn-nn-nn

Seuls les 6 premiers codes sont utilisés pour l'identification d'une entité. Le 7<sup>e</sup> renseigne sur le niveau de hiérarchie dans une division.

#### **Module 1= code1-code2 : de quoi s'agit-il ?**

- Code1 = code de l'époque et de la nature de la division (C11 = époque contemporaine, division administrative territoriale au sens INSEE : M21 = époque moderne, division fiscale Taille), sur 3 digits.
- Code2 = code d'un zonage (c'est un département, un comté, un bailliage, etc...), sur deux digits.

#### **Module 2 = code3-code4 : de qui s'agit-il ? (logique du code INSEE)**

- Code3 = code du niveau de référence, sur 3 digits, (le département pour la division administrative contemporaine, la généralité pour la division fiscale (Taille) de la période moderne, par exemple) dans laquelle sont codées les entités de niveaux inférieurs (les communes, les cantons, ou les paroisses fiscales ou les élections, par exemple).

Le territoire national est le niveau de référence (codé à « **000** ») pour les cas suivants :

- pour les niveaux de référence eux-mêmes (département, diocèse, parlement, intendance, généralité, etc.)
- pour les niveaux qui regroupent intégralement des niveaux de référence (les Régions qui regroupent des départements)
- pour les niveaux qui sont non couvrants (i.e que leur agrégation ne couvre pas la totalité du territoire national, les comtés, ou les parcs naturels, ou les communautés de communes par exemple)
- pour les niveaux qui ne s'appuient sur aucune limite administrative de base ou de référence (les régions géographiques)
- Code4 = numéro de l'entité du niveau inférieur, sur 4 digits, soit rapporté au territoire national (la Région, la Généralité, le Parlement, etc.), soit au niveau de référence (la commune ou le canton dans le département, le doyenné ou l'archidiaconé dans le diocèse, etc.).

La façon de **numéroter l'index (le code 4)** est celle de l'INSEE – et ici il faut bien comprendre le rôle que peuvent jouer les chef-lieux et les échanges de chef-lieux entre communes. En effet, on n'incrémente l'index des communes que lors d'une Création de commune (par une opération de distraction de parcelles sur d'autres communes). En revanche, dans le cas de l'absorption des communes, la commune absorbée prend l'index de la commune absorbante, et la commune absorbante :

- soit garde son index
- si un transfert de chef lieu a eu lieu, elle prend l'index de la commune où est situé le chef-lieu.

En cas de rétablissement, l'index de la commune rétablie est celui qu'elle avait avant son absorption.

Logiquement, lorsque l'unité change d'unité supérieure de référence, elle change donc de Code3. Ainsi une commune peut changer de canton ou d'arrondissement sans changer de code3. Seulement le passage dans un autre département change son code3.

### **Module 3 : code5-code6 = de quelle mutation s'agit-il ?**

- Code5 = codification des mutations appuyée sur celles retenues par l'Insee (cf. tableau 4), changement de nom, fusion, scission, échanges, transfert, etc. Les motifs de mutation, code 5, sont codés suivant la nomenclature proposée par l'INSEE<sup>9</sup>, afin d'assurer une compatibilité et comparabilité avec les informations que délivre l'INSEE sur l'histoire des communes. Seuls les entrées en rouge sont nouvelles par rapport aux codes que propose l'INSEE.
- Code6 = incrémentation des entrées dans la base – ou nombre de changements dans la même nomenclature. Lorsqu'une entité change de nomenclature, les modifications sont ré-indexées depuis 0.

Non affiché dans le code historique Géonoménclature.

- Code7 = niveau du zonage dans la hiérarchie de la nomenclature. Ce code, susceptible de changer dans le cas où de nouveaux zonages sont introduits dans une nomenclature ne peut pas être intégré au code historique. En effet, le code historique doit servir à indexer des documents, et ne doit pas être modifié.

---

<sup>9</sup> [http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc\\_variables.htm#mod](http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc_variables.htm#mod)

code	nom	Rq
00	1ere entrée dans la base	Non INSEE
01	Changement de nom	Non INSEE
02	Creation	Non INSEE
03	Suppression	Non INSEE
04	Changement d'appartenance	Non INSEE
05	Transfert de centre (chef-lieu)	Non INSEE
06	Transfert de parcelles	Non INSEE
07	Affectation à une nouvelle époque et/ou nouvelle division	Non INSEE
10	Changement de nom	
11	Changement de nom dû à une fusion (simple ou association)	
12	Changement de nom dû à un rétablissement	
13	Changement de nom dû au changement de nom du chef-lieu	
14	Changement de nom dû au transfert du chef-lieu	
20	Création	
21	Rétablissement	
22	Commune ayant donné des parcelles pour la création	
23	Commune se séparant	
24	Création d'une fraction cantonale	
30	Suppression	
31	Fusion	
32	Commune absorbante de la fusion	
33	Fusion - association	
34	Commune absorbante de la fusion-association	
35	Fusion-association se transformant en fusion simple	
36	Commune-pôle de la fusion-association qui s'est transformée en fusion simple	
37	Suppression de la fraction cantonale	
39	Commune de rattachement	
40	Changement de région	
41	Changement de département	
42	Changement d'arrondissement	
43	Changement de canton	
44	Changement d'appartenance (autres subdivisions)	Non INSEE
50	Transfert de chef-lieu de commune	
51	Transfert de chef-lieu de canton	
52	Transfert de chef-lieu d'arrondissement	
53	Transfert de chef-lieu de département	
54	Transfert de chef-lieu de région	
55	Transfert de chef-lieu (autres subdivisions)	Non INSEE
56	Ancien chef-lieu	Non INSEE
57	Nouveau chef-lieu	Non INSEE
60	Cession de parcelles avec incidence démographique	
61	Cession de parcelles sans incidence démographique	
62	Réception de parcelles avec incidence démographique	
63	Réception de parcelles sans incidence démographique	
64	Commune ancienne transférée	Non INSEE
70	Changement de code, affectation à une nouvelle époque et/ou nouvelle division	Non INSEE

Tableau 33. Liste des codes utilisés pour renseigner les motifs de mutation (événements) des entités.

### 3.3. Conception de la base de données

Outre les besoins qui viennent d'être exprimés, il s'agit également de tenir compte de **la nature imparfaite des données historiques** : elles peuvent être incomplètes (certains recensements manquent), imprécises et incertaines (une date de changement est approximée par un intervalle de plusieurs années). Ces informations sont également généralement caractérisées par des **métadonnées** qui renseignent sur les sources (bibliographiques, iconographiques ou autres) ayant servi à les établir ou les causes des lacunes. Ces informations, qui procèdent parfois de multiples conjectures peuvent être remises en cause et faire l'objet de **mise à jour**.

Nous proposons dans les paragraphes suivants une discussion concernant d'une part la nature des éléments à modéliser et leurs conséquences sur la modélisation dans un système de gestion de base de données, et, d'autre part, une proposition à même de répondre aux questions que pose la gestion sur le long terme de données imparfaites.

#### 3.3.1. Première hypothèse

La première hypothèse de conception postule qu'une entité territoriale n'a pas de vie, ni d'identité propre. Cette conception permet de s'abstraire des problèmes d'identification (reconnaissance d'une même identité territoriale à travers ces différentes versions). Une entité est alors définie **et identifiée** par les attributs suivants :

- Le toponyme
- Le territoire
- Le statut
- L'appartenance (ou lien hiérarchique) à une autre entité
- Le chef-lieu

Dès lors qu'un de ces attributs changent, une nouvelle entité territoriale est créée, peu importe que ce soit la « même » dans un sens commun : par exemple, si Lutèce devient Paris, on considère que deux entités territoriales se succèdent.

L'entité territoriale peut être une commune, une paroisse, un département, un canton, un arrondissement, etc., existant ou ayant existé dans les différentes subdivisions qui ont été définies. Ce modèle doit s'attacher à créer des liens de filiations entre les versions d'entités à travers un système de mutations qui lient entre elles des entités ayant précédé et succédé à un événement. Par exemple, Lutèce et Paris sont deux versions d'entités liées par un événement (une mutation) indiquant un changement de nom. Chaque version d'entité est datée, par un couple de date indiquant sa plage de validité (valide à partir de « debut », et modifiée en « fin »).

Elle possède également des attributs propres à sa version :

- sa surface (qui ne peut pas être calculée à partir de la géométrie : c'est une surface administrative) en hectares,
- son altitude maximale et minimale en mètres,
- un index signalant l'index de l'unité dans son groupe de référence,
- un compteur précisant le numéro de version de l'entité,
- un code historique.

Chaque version d'entité possède un code historique unique, comme décrit dans le paragraphe 3.2. Cependant, (et c'est peut-être contradictoire avec le déni de la notion de vie), une version d'entité peut être rétablie, et lorsqu'elle est rétablie, elle retrouve son ancien code. Afin de respecter la spécification concernant les opérations de rétablissement, le compteur de version n'est pas

incrémenté en cas de rétablissement. Cet élément de spécification implique que la version d'entité soit associée non pas à une seule plage de validité, correspondant à la durée pendant laquelle la version de l'entité est stable (sans changement d'appartenance, de géométrie, de nom, de statut ni de chef-lieux), mais à plusieurs plages de validité.

Outre les entités territoriales, un type particulier d'entité désigné par le terme « **élément** » doit être considérés avec attention, parce que ce sont des entités infra-territoriales, sans surface. Les historiens en donnent la définition suivante : « Elément rattaché à une commune sans territoire défini et sans fonction administrative ».

- Ils n'ont pas de géométrie
- Leur entité supérieure est l'entité à laquelle ils sont rattachés.

Leur plage de validité est indéterminée : ils sont valides depuis leur moment d'apparition dans les sources à de nos jours (décision des historiens).

Chaque entité territoriale appartient à un niveau dans une unique subdivision (=nomenclature). L'entité possède un unique lien hiérarchique avec le niveau de référence : une unité n'appartient qu'à une seule entité du niveau de référence dans sa subdivision. Cette entité de référence n'est pas forcément son unité immédiatement supérieure : ainsi, la commune appartient à un canton, mais se réfère à un département. Chaque entité nouvellement créée dans le groupe d'entités se référant à la même entité de référence possède un index unique dans ce groupe. Si la même (au sens commun) entité appartient à un autre niveau de référence dans une autre subdivision, une autre entité est créée dans la base, avec le code correspondant à cette autre organisation. **Ceci interdit tout lien inter-subdivision.**

Dans la même subdivision, l'entité peut avoir d'autres liens de hiérarchie avec des niveaux intermédiaires (mais qui ne soient pas de référence) : par exemple, une commune appartient à un canton, et une communauté de communes. Il est nécessaire de modéliser une relation d'appartenance à plusieurs niveaux simultanément. Une relation simplifiée (une entité n'appartient qu'à une seule entité supérieure), comme dans l'exemple de la table ENTITE\_TERRITORIALE\_SUPERIEURE (Tableau 34), ne permet en effet que de modéliser une hiérarchie stricte et onto.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
identite_territoriale	integer	FK	Entité territoriale
identite_territoriale_liée	integer	FK	Entité supérieure d'un autre niveau

Tableau 34. La table ENTITE\_TERRITORIALE\_SUPERIEURE.

Par exemple, (Tableau 35), si la commune ‘c1’ appartient au canton ‘c12’, qui appartient à l’arrondissement ‘arr200’ qui appartient au département ‘dept01’, alors l’entité de référence de la commune ‘c1’ est le ‘dept01’.

Entité	Entité supérieure
Commune c1	Canton can12
Canton can12	Arrondissement arr200
Arrondissement arr200	Département dept01

Tableau 35. Exemple de hiérarchie stricte et onto.

Le problème est que généralement l'ensemble des divisions et subdivisions traitées ne sont ni strictes ni onto. Et ainsi la table ne peut pas permettre de trouver qu'elle est l'entité de référence. Par exemple (Tableau 36), la commune **c2 appartient à la communauté de communes epci1 qui appartient au département dept01**, mais **c2 appartient également au département dept02** (par le

biais de son canton et son arrondissement) : quelle est son entité de référence ?

Par ailleurs, il n'est pas évident de vérifier si une même entité appartient à deux entités de même niveau (comme **c1 appartient à epci1 et epci2**).

Entité	Entité supérieure
Commune c1	Communauté de commune epci1
Commune c1	Canton can12
Canton can12	Arrondissement arr200
Arrondissement arr200	Département dept01
<b>Commune c2</b>	Communauté de commune epci1
Communauté de commune epci1	<b>Département dept01</b>
<b>Commune c2</b>	Canton22
Commune c2	Arrondissement arr222
Arrondissement arr222	<b>Département dept02</b>
Commune c1	Communauté de commune epci2

Tableau 36. Exemple de hiérarchie non stricte et non onto.

En complétant la spécification de cette table par **l'insertion obligatoire des appartenances à tous les niveaux** pour chaque entité (Tableau 37), il serait possible de lire directement l'entité de référence (voir le Tableau 38) : celle dont le niveau est celui du niveau de référence (ou celle qui se référence elle-même comme entité de référence). Cette précision permettrait également de gérer des cas d'exception comme des communautés de communes à cheval sur plusieurs départements (c'est fréquent) et déclarées comme faisant partie d'un département, tandis que certaines de leur communes appartiennent à d'autres départements. Egalement, il serait possible de voir directement qu'une entité paie son impôt à deux entités de même niveau supérieure durant la même période.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
identite territoriale	integer	FK	Entité territoriale
identite territoriale liée	integer	FK	Entité supérieure d'un autre niveau
niveau	Integer	FK sur NIVEAU	Identifiant du niveau de l'entité
Niveau_sup	Integer	FK sur NIVEAU	Identifiant du niveau de l'entité supérieure

Tableau 37. La table ENTITE\_TERRITORIALE\_SUPERIEURE.

Entité	Entité supérieure	/Niveau	/Niveau supérieur
Commune c1	Communauté de commune epci1	5	4
Commune c1	Canton can12	5	3
Canton can12	Arrondissement arr200	3	2
Arrondissement arr200	Département dept01	2	1
<b>Commune c2</b>	<b>Communauté de commune epci1</b>	5	4
Communauté de commune epci1	Département dept01	4	1
Commune c2	Canton can22	5	3
Canton can22	Arrondissement arr222	3	2
Arrondissement arr222	Département dept02	2	1
Commune c1	Communauté de commune epci2	5	4
Commune c1	Arrondissement arr200	5	2
<b>Commune c1</b>	<b>Département dept01</b>	5	1
Canton can12	Département dept01	3	1
Commune c2	Arrondissement arr222	5	2
<b>Commune c2</b>	<b>Département dept02</b>	5	1
Canton can22	Arrondissement arr222	3	2
Canton can22	Département dept02	3	1

Tableau 38. Exemple de hiérarchie non stricte et non onto, avec spécification de toutes les appartenances (en grisé).

Cette hypothèse, où les attributs ne sont pas versionnés, mais seulement les entités, présente néanmoins des défauts majeurs.

Dès lors que le toponyme, la géométrie, le statut, l'appartenance (ou lien hiérarchique) à une autre entité et le chef-lieu d'une entité change, il faut **duplicer** entièrement l'entité, et ses attributs.

- Premièrement, ceci est coûteux en terme d'espace (en particulier lorsqu'une géométrie (un multi-polygone) est composée de 200 points en moyenne, soit 3 Ko).
- Ensuite, les requêtes sont difficiles : les recensements associés à chaque version d'entité deviennent difficiles à récupérer pour un ensemble d'entités ayant évolué au cours du temps, puisqu'il faut suivre les liens de filiation entre les différentes versions d'entité pour récupérer les données.
- Enfin, la complexité des modifications en cascade qu'un tel paradigme peut entraîner pose problème. Ainsi, par exemple, en cas de changement de nom en date D2, un canton can1 est dupliqué et devient le canton can2 par exemple. Ce canton gouvernait les communes c1, c2, c3 et c4. Ces communes n'ont pas changé de canton (réellement) mais par un artifice, elles ont maintenant un autre canton supérieur (can2). La conséquence logique serait donc la duplication de c1, c2, c3 et c4 et donc de leurs attributs.

Ceci nous a donc conduit à réviser cette hypothèse de conception.

### 3.3.1. Seconde hypothèse

Dans cette seconde hypothèse de conception, les entités ont une vie (et donc un identifiant système unique au cours du temps), avec des attributs versionnés au cours du temps. Des classes d'associations datées (qui se traduisent sous la forme de table de jointure) modélisent l'évolution de liens entre les éléments du modèle. Ces classes d'association existent entre :

- le chef-lieu et l'entité qui est chef-lieu,
- l'entité et son entité supérieure,
- l'entité est un zonage dont elle fait partie.

La Figure 8 présente une vue générale du modèle finalement adopté.

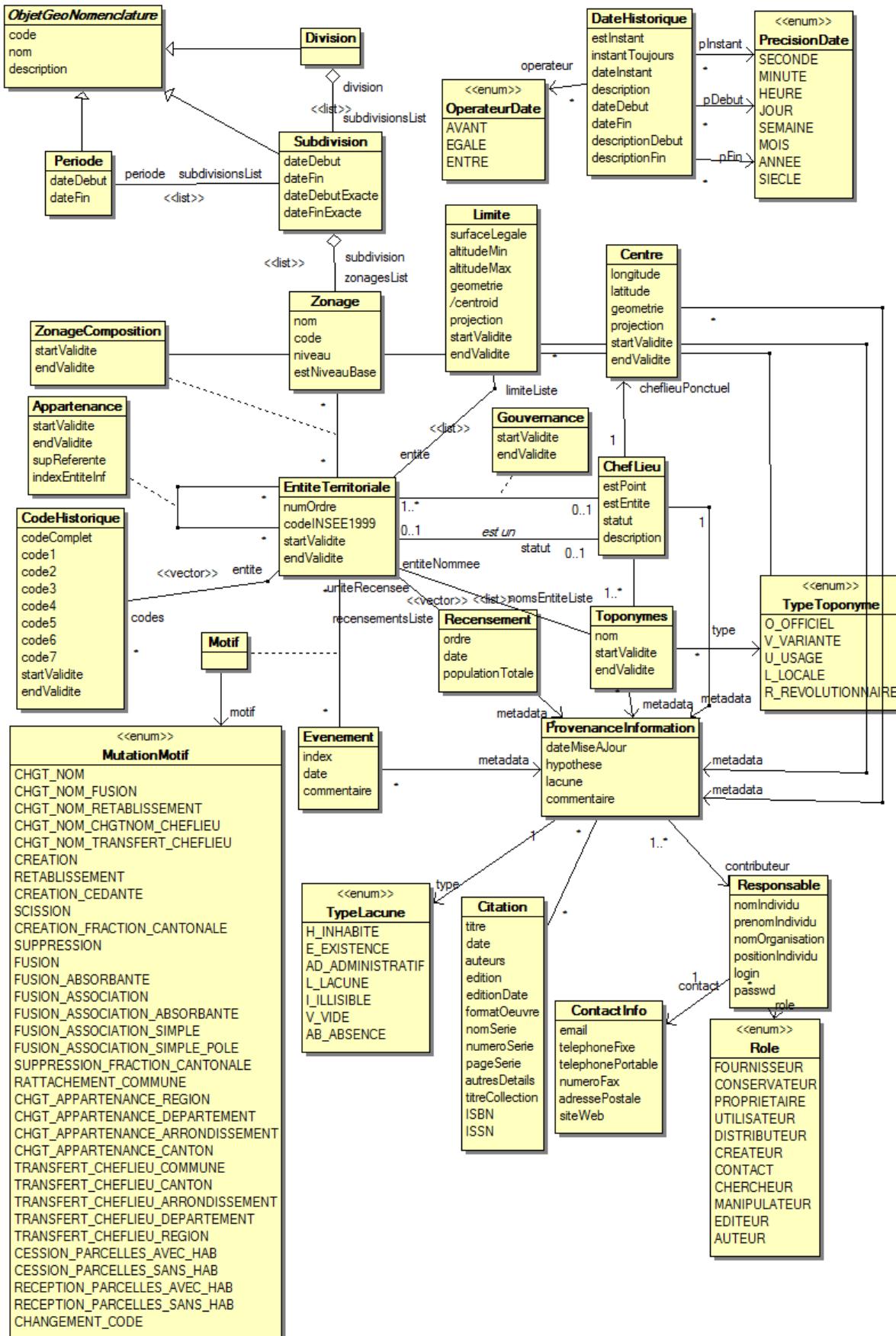


Figure 8. Modèle des données historiques dans GeoPeople.

Ce modèle répond à plusieurs préoccupations :

- l'enregistrement de nomenclatures autres que celles de l'INSEE, non forcément strictes et onto,
- la gestion de données imparfaites, qui se traduit par l'introduction de métadonnées, ainsi que d'un modèle de datation des éléments original,
- l'historicisation des mises à jour de données,
- la compatibilité avec d'autres bases (celle de l'INSEE et celle de Cassini),
- une optimisation des performances et de l'espace mémoire utilisé.

Les sections qui suivent proposent une description détaillée de chaque élément de ce modèle, en justifiant en détail les choix faits.

### 3.3.2. Gestion des nomenclatures

GeoNomenclature propose des hiérarchies de nomenclature, classées par des époques. Les nomenclatures (divisions, sous-divisions et zonages) proposées et leur datation ne sont pas éditables par un utilisateur non expert. Elles figent la base de données et constituent sa structure : ce sont des constantes de la base. Elles n'ont pas encore été toutes spécifiées (toutes les sous-divisions ne sont pas énumérées). Il faut se référer au document « Géonoménclature historique des lieux habités » pour trouver la spécification complète de ces données, déposée aux Archives de France.

Division, Subdivision, et Periode héritent de l'objet ObjetGeoNomenclature qui définit trois champs de type 'chaînes de caractères' : un nom, une description et une valeur. La valeur est un code alphanumérique de longueur 1. La concaténation de ces valeurs forme le code1 du code historique. Pour l'instant, très souvent, la description est une copie du nom, comme le montre le Tableau 39.

code	nom	description
0	Division géographique	Division géographique
1	Division administrative	Division administrative
2	Division financière	Division financière
3	Division religieuse	Division religieuse
4	Division judiciaire	Division judiciaire civile et pénale
6	Division académique	Division académique
C	Division autres	...

Tableau 39. Liste des divisions prédéfinies.

Trois époques (périodes) sont proposées : Moyen-Âge, Epoque moderne, Epoque contemporaine. Une période est associée à un ensemble de Subdivisions (ou nomenclatures). La période de validité des entités territoriales n'est cependant pas limitée par ces époques : une même entité peut appartenir successivement à deux nomenclatures d'époques différentes.

code	date_debut	date_fin	nom	description
F	476	1492	Moyen Âge	époque médiévale : de la chute de l'empire romain (476) jusqu'à la prise de Grenade (1492)
M	1492	1789	époque moderne	époque moderne : de la prise de Grenade (1492) à la création des départements (février 1790)
C	1790		époque contemporaine	époque contemporaine : de la création des départements (février 1790) à aujourd'hui

Tableau 40. Périodes.

Les sous-divisions (ou subdivisions) sont associées à des plages de validités, qui doivent être incluses dans les époques. Les dates des subdivisions ne sont pas toutes connues/renseignées. Par exemple, pour les sous-divisions dont la liste des zonages a été établie, seule la sous-division Département (1801) de l'époque contemporaine possède une date de début qui est 1801, et la date de fin doit indiquer qu'elle reste valide de nos jours. Par ailleurs, en plus d'une plage de validité spécifiée par des dates de début et de fin, des dates de début et de fin exactes, sensées être plus précises sont spécifiées. Elles indiqueraient la date d'un événement historique marquant le début ou la fin de la subdivision avec une granularité plus fine (au jour près par exemple) que les autres dates. Par exemple, la date exacte de l'acte ou la loi fondant le début de la sous-division Département 1801 est le 17 Février 1800.

Sub division	période	division	code	nom	début	début exact	fin	fin exacte
1	C	0	0	Sous-division géographique	1790	1790-01-01	*	*
2	C	1	0	Division intercommunale	1790	1790-01-01	*	*
4	C	1	2	Division de la France d'outre-mer	1801	1801-01-01	*	*
6	C	1	9	Pays et Territoires étrangers	1801	1801-01-01	*	*
7	C	3	1	Culte catholique	1790	1790-01-01	*	*
8	C	3	2	Culte réformé	1790	1790-01-01	*	*
9	C	3	3	Culte israélite	1790	1790-01-01	*	*
3	C	1	1	Division de l'administration territoriale métropolitaine 1801	1801	1800-02-17	*	*
5	C	1	3	Division administrative (1790-1800)	1790	1790-01-01	1801	1800-02-17

\* toujours valide aujourd'hui

Tableau 41. Extrait des subdivisions définies pour GeoNomenclature.

Un zonage est décrit par un code (de type entier) qui sert à renseigner le code2 du code historique, un nom, et un niveau (niveau dans la hiérarchie). L'information « niveau\_base » signale si un niveau est un niveau plancher pour la subdivision.

Nous proposons la définition suivante pour **niveau plancher ou niveau de base** :

- Un niveau de base forme une partition complète et sans recouvrement du territoire
- Il est possible de définir le territoire des entités de niveau supérieur par agrégation des limites des territoires du niveau de base.

En annexe, page 71, la liste complète des **zonages** définis dans GeoNomenclature est présentée dans le Tableau 77.

### 3.3.3. La gestion du temps de validité

Afin de tenir compte de façon plus rigoureuse et systématique de l'incertitude pouvant exister sur la datation des éléments, un nouveau modèle de datation de tous éléments est introduit. En effet, dans Cassini, pour de nombreux éléments, il a été observé que la date de début ou bien de fin d'un élément (toponyme, ou événement par exemple) n'était pas forcément connue avec certitude, et que pour faire face à ce problème, les solutions n'étaient pas toujours harmonisées pour tous ces éléments, ni optimales en terme d'espace mémoire ou de requête.

En règle général, les champs « date\_debut » et « date\_fin » doivent absolument être du type *date* (voir la page de documentation de postgres <http://www.postgresql.org/docs/8.0/static/datatype-datetimehtml>). En effet, si ces dates doivent servir à des validations de contraintes, ou à des opérations de tri, il n'est pas acceptable d'utiliser un champ de type chaîne de caractères dans le

modèle. Si à l'affichage, la date doit être présentée dans un certain format, il existe quantité de bibliothèques permettant de formatter la date de la bonne façon. Cette remarque s'applique à tous les champs date de la base de données. Enfin, il s'agit d'introduire une distinction claire entre :

- Date inconnue (c'est une lacune)
- Date « toujours valide »

La solution de prendre une date dans le futur (comme 2500) ne nous semble pas conceptuellement très élégante, ni non plus d'utiliser 'NOW' comme marqueur de validité [Clifford et al., 1997], ou d'utiliser le mot de clé d'un SGBD (comme '*infinity*' dans le cas de PostgreSQL) car cette solution n'est jamais absolument portable d'un système à un autre.

Nous proposons d'utiliser un nouveau type complexe de date (illustré par la Figure 9) qui définit :

- Si la date est connue comme un instant ou par un intervalle
- Si la date précède, est exactement valide à, ou succède à l'instant défini, ou bien appartient à la période définie.

L'instant ou l'intervalle peuvent être décrits :

- soit par des dates de type Date, avec une précision variable (jour, mois, année, siècle),
- soit par un champs textuel qui définit un jour (par exemple, 'Sacre de Napoleon premier') ou une période ('L'Empire').

Dans le cas où les dates sont définies uniquement par ces champs textuels, les comparaisons entre dates nécessitent d'ordonner ces époques et instants décrits<sup>10</sup>. Les opérateurs définis (Tableau 42) peuvent être étendus, mais nous n'avons proposé que ceux utilisés dans la base de cassini.

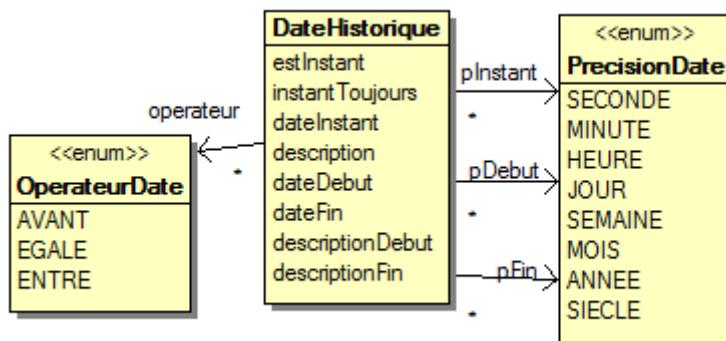


Figure 9. Le type DateHistorique.

Ce nouveau type de date, identifié uniquement par un entier, est référencé par tous les champs de datation du modèle : l'index étant posé sur cet identifiant défini comme clé primaire de l'objet DateHistorique, les performances de recherche sont très bonnes, et par ailleurs, une gain d'espace mémoire important est ainsi réalisé dans la mesure où beaucoup de champs référencent les mêmes dates.

DateHistorique est initialisé avec deux Dates particulières : la date « toujours valide », id = 1, qui signifie que les éléments associés restent encore valides aujourd'hui, et la date « inconnue », id=2, spécifiant expressément que cette date n'est pas connue.

IDOPERATEUR	VALEUR	NOM	DESCRIPTION
-1	$\leq$	avant	L'élément daté est valide avant la date
0	=	égale	L'élément daté est valide à cette date exactement
1000	[P]	entre	La date de l'élément se situe durant cette période, entre le début de la période et la fin de la période.

Tableau 42. Exemples d'opérateurs définis pour TYPE\_OPERATEUR\_DATE.

<sup>10</sup> En réalité, cela ne se produit pas. Le texte vient en complément des dates uniquement.

Afin de faciliter les opérations de tri, de recherche et d'affichage des données associées à ces dates, nous avons également définis des procédures stockées en PL-SQL (leur pré-compilation au sein du SGBD garantie leur performance). Ainsi les trois procédures suivantes permettent de rechercher ou créer rapidement des dates historiques à partir des données spécifiées dans Cassini (souvent des dates spécifiées sous forme d'entier, avec l'année comme granularité).

**history\_getDateHistorique\_instantYEAR(x entier) : entier**

- paramètre d'entrée : une année (exprimée dans le calendrier grégorien, comme un entier : 1990 par exemple)
- renvoie l'identifiant de la date historique correspondant dans la base
- si l'entrée n'existe pas pour cette date, elle est créée

**history\_getDateHistorique\_periodeYEAR(x entier, y entier) : entier**

- paramètre d'entrée : deux années (exprimées dans le calendrier grégorien, comme des entiers : 1990 et 1999 par exemple), qui représentent un intervalle temporel [x, y]
- renvoie l'identifiant de la date historique correspondant dans la base
- si l'entrée n'existe pas pour cette date, elle est créée

**history\_getDateHistorique\_instantBeforeYEAR(x entier) : entier**

- paramètre d'entrée : une année (exprimée dans le calendrier grégorien, comme un entier : 1990 par exemple)
- renvoie l'identifiant de la date historique correspondant ('Avant 1990' par exemple) dans la base;
- si l'entrée n'existe pas pour cette date, elle est créée

La procédure **history\_afficheDateHistorique\_asDate** est utilisée pour faire des recherches triées par date sur tous les champs datés de la base, ou bien des requêtes impliquant une comparaison sur les champs dates :

```
select ... order by history_afficheDateHistorique_asDate(champs_date)
select ... where history_afficheDateHistorique_asDate(champs_date1) < '17/02/1801'
```

**history\_afficheDateHistorique\_asDate(x entier) : date**

- paramètre d'entrée : l'identifiant d'une date historique
- renvoie la date (ou la borne supérieure) de l'intervalle correspondant ; pour 'avant 1990', renvoie '1990' ; pour entre 1990 et 1999, renvoie 1999

La procédure **history\_afficheDateHistorique\_asString** est utile pour l'affichage et la compréhension des données.

**history\_afficheDateHistorique\_asString(x entier) : Chaîne de caractères**

- paramètre d'entrée : l'identifiant d'une date historique
- renvoie une chaîne de caractères correspondant à cette date comme 'avant 1990', 'entre 1990-01-01 et 1999-01-01', ou 'toujours valide aujourd'hui' ou 'inconnue'

Les deux procédures suivantes permettent de trier des dates historiques.

**history\_isDateAvant(x entier, y entier) : booléen**

- paramètre d'entrée : identifiant de deux dates historiques
- renvoie vrai si Y est avant ou touche le début de X (si Y<= X alors vrai), faux sinon et faux si x vaut 1 (toujours valide)

**history\_isDatePendant(x entier, y entier) : booléen**

- paramètre d'entrée : identifiant de deux dates historiques

- renvoie vrai si  $y.\text{binf} \geq x.\text{binf}$  and  $y.\text{bsup} \leq x.\text{bsup}$  et faux si  $x$  vaut 1 (toujours valide)

### 3.3.4. Les métadonnées de provenance.

Le schéma intègre une connaissance sur les contributeurs, les sources de données utilisées, et une typologie des lacunes, désignées et référencées par les « **metadata** ». Les relations permettent de retracer la provenance de l'information et de la qualifier. Le modèle s'inspire directement de la norme ISO 19115 pour les métadonnées géographiques.

A chaque élément du modèle est associé une métadonnée « ProvenanceInformation » qui décrit la *date de mise à jour* de l'information (dateMiseAJour), c'est-à-dire la date d'enregistrement de l'information dans le système. Ceci répond au besoin de mise à jour de ces informations qui font l'objet de recherches, et dont les conclusions peuvent être révisées à la lumière de nouvelles sources. Un champ « hypothèse » de type booléen précise si l'information est certaine, ou résulte d'une estimation de l'auteur, comme c'est le cas pour les limites anciennes des communes. Un commentaire libre peut être inséré dans le champ « commentaire ».

Le champ « lacune » de type booléen permet de préciser si cette information est lacunaire, et le *type de lacune* est précisé. Pour la typologie de lacune (Tableau 43), nous sommes repartis de celles établie par EHESS pour les recensements, en étendant à des symboles observés dans la base mais non documentés.

code	nom	definition	code_cassini
0	pas de lacune	information complète - pas de lacune	
1	Inhabité	Entité inhabitée	0 hab.
2	Inexistence	Entité n'existant pas à cette date	...
3	Administratif	Entité recensée avec une autre	adm.
4	Lacune	Entité oubliée sur la publication du recensement	lac.
5	Illisible	Information illisible	ill.
6	Vide	Information non disponible actuellement	vide
7	Absence	Document (ou pages) disparu(es)	abs.
8	Null value	l'information est inconnue, et rien au sujet de la lacune	NULL
9	Autre lacune	Le code lacune n'était pas connu	xxx

Tableau 43. Typologie des types de lacune.

Par ailleurs, les métadonnées référencent qui sont les *auteurs* (ou contributeurs) de ces informations. Ils sont décrits par l'élément « Responsable » référencé par « ProvenanceInformation », via les attributs suivants :

- le nom de l'individu,
- le prénom de l'individu,
- le nom de l'organisation,
- la position de l'individu dans l'organisation,
- le login de l'individu sur le système d'information contenant les données,
- le mot de passe de l'individu sur le système d'information contenant les données.

Afin de suivre la norme ISO19115, mais de façon simplifiée, ce responsable est associé des informations de contact (l'élément ContactInfo) qui précise l'adresse de courrier électronique, le téléphone fixe et portable, le numéro de fax, le site web, et l'adresse postale du contact. Seule l'adresse de courrier électronique est obligatoire.

L'ensemble de ces informations doit permettre à terme de gérer cette nomenclature de façon collaborative, en autorisant des contributeurs multiples. En effet, un **système collaboratif** nous

semble un moyen pour rendre plus vivant le système d'information, et donc pérenne au final. Notre proposition est cependant limitée au point de vue de la granularité des métadonnées, qui ne s'attachent pas à tous les attributs des relations, mais seulement aux relations elles-mêmes.

### 3.3.1. La gestion du versionnement de l'information (temps de transaction)

Ayant prévu un système de mise à jour collaborative, nous proposons également une solution ad-hoc, mais encore embryonnaire pour l'historicisation des données, c'est-à-dire la conservation des données ayant fait l'objet de modifications dans la base de données. Pour cela, une relation « history version » doit permettre d'enregistrer les anciennes informations. Ceci permet de récupérer d'erreurs ou de mauvaises mises à jour, mais également de retrouver les traces d'informations qui auraient pu être intégrées à des études scientifiques avant d'être révisées.

Cette relation (Tableau 44) décrit la classe de l'objet modifié, son identifiant, sa valeur avant modification (sous la forme sérialisée d'une chaîne de caractères non limitée en longueur, quelque soit le type original de l'information), l'ordre de sa mise à jour, la date de mise à jour et les métadonnées associées à sa valeur précédente.

Champ	Type	Contrainte	Description
classe	character varying(255)	NOT NULL	Nom de la table modifiée, correspondant à la classe de l'objet
idobjet	integer	NOT NULL	Nom de la colonne modifiée, correspondant à l'attribut modifié
valeur	character varying	NOT NULL	Valeur de l'attribut modifié (casting vers une chaîne de caractères)
ordre_maj	integer	NOT NULL	Ordre de mise à jour dans le système pour cet objet (première, deuxième, troisième révision) qui peut affecter un ou plusieurs champs (attributs) de l'objet
date_maj	timestamp	without time zone	Date et heure de cette mise à jour (moment où l'élément est archivé)
metadata	integer	FK vers history_provenance_information (idprovenance)	

Tableau 44. Table VERSIONS.

### 3.3.1. Toponymes et chef-lieux

Nous rappelons que :

- Une entité a toujours au moins un toponyme, et ce toponyme peut-être daté, et sa datation est indépendante de celle de l'entité, et reste facultative.
- Une entité peut aussi avoir des variantes de toponyme, datées elles-aussi, comme le toponyme.

Les différentes versions (ou variantes) de toponymes associées à une entité sont issues de la consultation de sources historiques (actes, recensements, etc.) dont on a reconnu qu'elles désignaient une même entité sous différentes dénominations. En particulier, par exemple, sous la période du directoire, après la révolution française, un nombre important de communes ont été rebaptisées avec des noms révolutionnaires, supprimant les références aux noms de saints catholiques. Mais les deux types d'appellation, d'origine et révolutionnaire, ont souvent continué de co-exister dans les faits pendant une certaine durée. De plus, des erreurs ou des coquilles se sont

parfois glissées dans les documents administratifs et ainsi des versions à l'orthographe divergentes coexistent jusqu'à rectification dans les sources historiques pour une même entité. Ceci nous a conduit à établir une nomenclature des types de toponyme, (Tableau 45). On considérera qu'un nom est officiel quand il est retrouvé dans les Actes de cette commune, même si en réalité, il n'existe pas de nom déclaré officiellement.

code	nom	description
O	Officiel	Toponyme officiel
R	Forme révolutionnaire	Forme révolutionnaire (entre 1790 et 1801 souvent)
U	Dénomination d'usage	Dénomination d'usage pour les entités sous niveau de base (infra-communales)
V	Variante	Forme défectueuse d'un toponyme rencontrée alors qu'une forme officielle existe ou a existé.
L	Dénomination en langue locale	Dénomination en langue locale

Tableau 45. Types de toponyme.

Normalement toute entité territoriale du modèle possède au plus un chef-lieu en même temps. Egalemente, certaines entités territoriales, de par leur nature, ne semblent pas avoir de chef-lieu (si elles sont de statut inférieur au niveau de base par exemple, comme les « Elément rattaché à une commune sans territoire défini et sans fonction administrative » ou les « Elément rattaché à une paroisse »). Bien qu'un chef-lieu soit toujours rattaché à une entité territoriale, il peut posséder son propre toponyme<sup>11</sup>. Parfois ce chef-lieu n'est connu que par son toponyme. Concernant sa nature, il y a deux cas pour un chef-lieu « x » :

- « x » est un point, donc n'a pas de surface : il faut stocker la géométrie de ce centre sous la forme d'un Point. C'est le sens de l'association entre un ChefLieu et un « Centre ». Dans ce cas, le booléen « estPoint » prend la valeur vraie.
- « x » est une autre entité : il faut référencer cette entité. C'est le sens de l'association « est un » entre un ChefLieu et une EntiteTerritoriale. Dans ce cas, le booléen « estEntite » prend la valeur vraie. La période de validité du ChefLieu se réfère à la durée pendant laquelle l'entité est chef-lieu.

Les deux cas ne sont pas exclusifs : x peut à la fois être connu par des coordonnées géographiques, et comme entité. Exemple : la commune de Grenoble est chef-lieu de l'Isère, mais son centre est repéré par les coordonnées de la préfecture.

Le rôle de Chef-lieu est de gouverner des entités. Le terme de Chef-lieu est propre à la nomenclature INSEE : chef-lieu de canton, chef-lieu d'arrondissement (ou sous-préfecture), chef-lieu de département (ou préfecture). Il aurait pu s'appeler de façon plus générique « Centre de gouvernance ». Il existe deux façons de préciser celui-ci :

- soit on spécifie qu'elles sont les entités de même niveau sous le contrôle du chef-lieu, avec la période de validité de cette gouvernance,
- soit on spécifie si possible l'entité de niveau supérieur qui regroupe toutes les entités sous contrôle, avec également la période de validité de cette gouvernance.

Ainsi, une commune sera chef-lieu d'un canton (relation 1 :1 entre deux entités de niveau différents) ou bien une commune sera chef-lieu d'un groupe de communes appartenant à ce canton (relation 1 :n entre deux entités de même niveau). Par rapport au contenu des sources historiques, il est souvent plus facile d'utiliser la première façon de noter cette relation, que nous privilégions. La classe d'association « Gouvernance » permet de spécifier la durée de relation de gouvernance existant entre un chef-lieu et les entités gouvernées.

<sup>11</sup> Même si il possède souvent le même toponyme que l'entité territoriale dont il est chef-lieu, le toponyme du chef-lieu peut être différent de celui de l'entité territoriale.

### 3.4. Compatibilité avec d'autres bases existantes

La compatibilité avec les bases de données INSEE est assurée via la reprise du code INSEE (en 1999) des communes dans le champ **CODE\_INSEE\_1999** de l'entité territoriale, lorsque l'entité territoriale est une commune. Ce code permet un croisement avec les informations existantes dans la base de données de l'INSEE publiée sur le Web, et la détection d'éventuels conflits<sup>12</sup>.

Concernant les codes des motifs de mutation, nous utilisons un motif codé suivant la nomenclature proposée par l'INSEE<sup>13</sup>, cf. paragraphe 3.2 page 32, Tableau 33. En effet, le **motif** qualifie ce qui arrive à chaque entité actrice d'un événement (elle est absorbée (30), absorbante (32), créée (20), elle cède des parcelles pour la création (22) ou simplement pour un échange de territoire avec impact démographique (60), etc. Le schéma Figure 10 illustre les principaux événements traités dans la base.

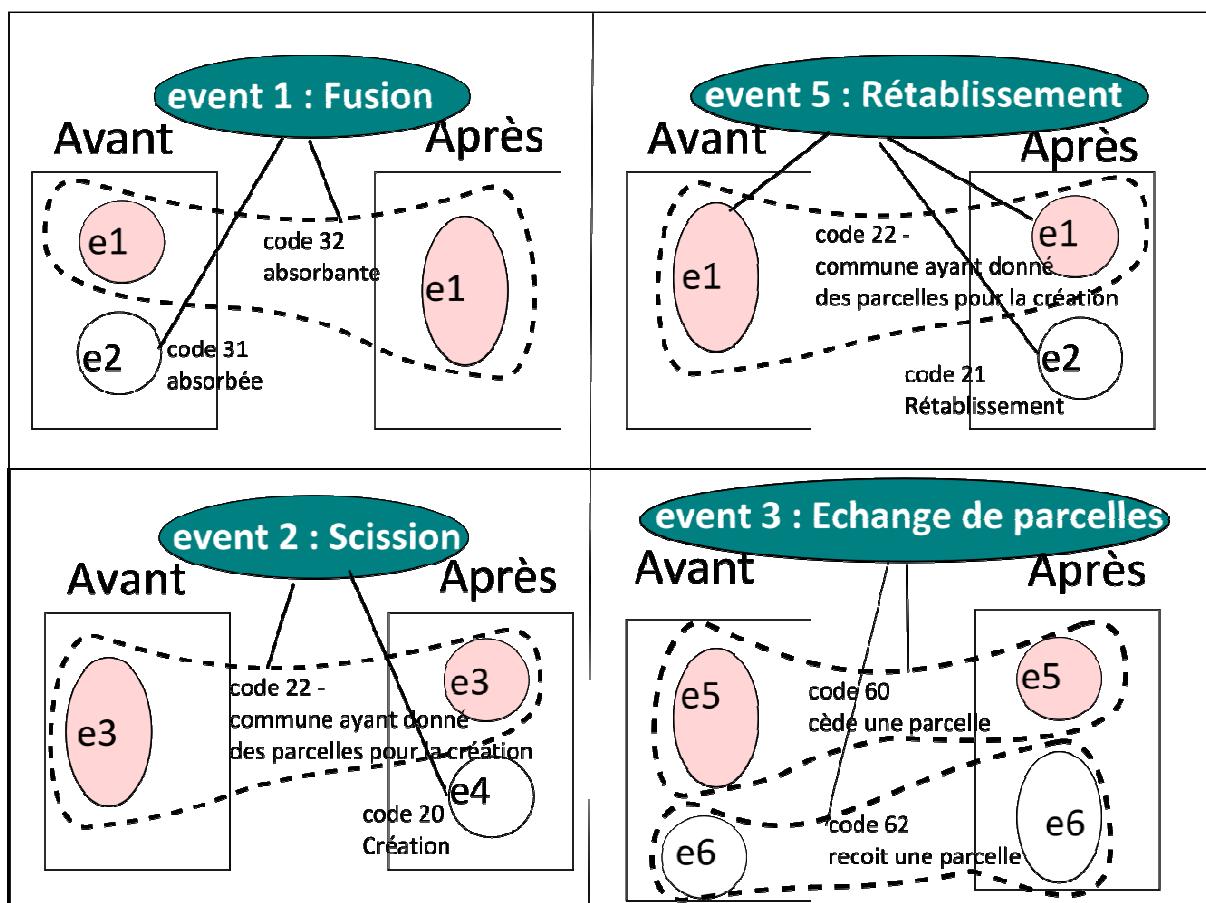


Figure 10. Événements territoriaux les plus fréquents.

Enfin, lorsque l'entité territoriale est issue d'une entité territoriale décrite dans la base de données Cassini, nous avons souhaité également conservé son numéro d'ordre (*num\_ordre*) qui l'identifiant uniquement sous cassini. Ainsi, chaque entité est décrit par :

- un identifiant système unique de type entier, propre à GeoPeople,
- le code insee 1999 long de l'entité si c'est une commune, donné par l'INSEE,
- le numéro d'ordre (*num\_ordre*) correspondant dans la base de données Cassini.

De même les événements possèdent un identifiant système unique, mais également l'identifiant de l'événement exporté depuis la base de données Cassini (*index\_evenement\_cassini*).

<sup>12</sup> [http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc\\_fhistor.htm](http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc_fhistor.htm)

<sup>13</sup> [http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc\\_variables.htm#mod](http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/cog/doc_variables.htm#mod)

## 4. Implémentation et migration de la base de données

Nous décrivons ici les spécifications relatives à certains champs de la base de données GeoNomenclature, ainsi que les modalités de la migration. Toutes les tables décrites sont précédées du suffixe « history ».

Le SGBD choisi pour supporter cette base de données relationnelle est Postgres (<http://www.postgresql.org/>), un outil libre, qui propose également une cartouche spatiale PostGIS (<http://www.postgis.org/>) permettant de gérer des données géographiques. Le SGBD Postgres est conforme au langage standardisé SQL, multi-plateforme (Windows, Linux, Mac), labellisée ACID (les transactions sont protégées), et propose un nombre important d'interfaces programmatiques natives (Java, C/C++, Python, etc.). Le niveau de développement de la communauté open-source autour de ces deux produits assure la pérennité de ce choix.

Nous avons essayé de proposer une implémentation du modèle la plus optimale possible. Dans cette optique, les contraintes d'intégrité ont bien été spécifiées. En effet, une clé primaire (notée « PK »), déclenche à la fois une contrainte d'unicité et de non nullité, mais également la mise en place d'un index pour les recherches sur ce champ (ou ensemble de champs) sous PostgreSQL. Egalement, la spécification des clés étrangères, (notées « FK ») permet de maintenir facilement la cohérence de la base de données en cas de suppression d'éléments, et génère automatiquement un index sur le champ référencé.

### 4.1. Description des relations

#### 4.1.1. Les relations temporelles

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idate	serial	PK	Id
est_instant	boolean	NOT NULL	Vrai si la date est un instant, faux si intervalle (entre date_debut et date_fin).
instant_toujours	Boolean	NOT NULL	Vaut vrai si la date (qui est un instant) est aujourd'hui.
operateur	character varying(255)	FK vers OPERATEUR_DATE, '=' par défaut.	Operateur à utiliser avec la date (est exactement, avant, après, pendant, en dehors de, etc.) défini dans OPERATEUR_DATE
date_instant	date		Si la date est définie comme un instant, ce champ la décrit avec la précision qui convient à l'utilisateur.
description	character varying(255)	Default NULL	Description textuelle de l'instant ou de la période.
date_debut	date		Si la date est définie par une période, cette date décrit le début de la période, avec la précision qui convient à l'utilisateur.
date_fin	date		Si la date est définie par une période, cette date décrit la fin de la période, avec

			le précision qui convient à l'utilisateur.
description_debut	character varying(255)	Default NULL	Description textuelle de l'instant débutant la période.
description_fin	character varying(255)	Default NULL	Description textuelle de l'instant terminant la période.
precision_instant	integer	Default 1, FK sur la	Precision de la date - si précision à l'année, jour = 01 et mois = 01 - si précision au mois, jour = 01
precision_debut	integer	Default 1	Precision de la date - si précision à l'année, jour = 01 et mois = 01 - si précision au mois, jour = 01
precision_fin	integer	Default 1	

Tableau 46. La table DATE\_HISTORIQUE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idoperateur	integer	PK	Id
valeur	character varying	NOT NULL	Un symbole ou une chaîne courte
nom	character varying(255)	NOT NULL	Nom de l'opérateur : avant, égale, entre, ...
description	character varying(255)		Description de l'opérateur

Tableau 47. La table OPERATEUR\_DATE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
code	integer	PK	Un entier, identifiant cette précision : 1 pour seconde, 3 pour jour, 6 pour année.
nom	character varying(255)	NOT NULL	nom de ce niveau de précision
description	character varying(255)		description de ce niveau de précision

Tableau 48. La table PRECISION\_DATE.

#### 4.1.1. Les tables de métadonnées

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
code	integer	PK	code du rôle
nom	character varying(255)	NOT NULL	Nom court en français
definition	character varying(255)		Définition en français
Nom_en	character varying(255)	NOT NULL	Nom court en anglais
Definition_en	character varying(255)		Définition en anglais

Tableau 49. La table ROLE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idcontact	integer	UNIQUE	identifiant système unique
email	character varying(255)	PK	adresse électronique (email), obligatoire, d'un acteur ou contributeur
telephone_fixe	character varying(20)		Numéro de téléphone fixe (format international) d'un acteur ou contributeur
telephone_portable	character varying(20)	NOT NULL	Numéro de téléphone portable (format international) d'un acteur ou contributeur
telephone_fax	character varying(20)	NOT NULL	Numéro de fax (format international) d'un acteur ou contributeur
adresse_postale	character varying	NOT NULL	adresse postale, définie suivant le format

			suivant : "numéro=;batiment=;rue=;etat=;ville=;codePostal=;pays="
site_web	character varying(255)	NOT NULL	URL du site Web de l'organisme ou de la personne

Tableau 50. La table CONTACTINFO.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idresponsable	integer	UNIQUE	identifiant système unique.
idcontact		FK sur CONTACT_IN FO	informations nécessaires pour contacter un contributeur.
idrole		FK sur ROLE	role du contributeur.
prenom_individu	character varying(20)	NOT NULL	prénom du contributeur.
nom_individu	character varying(20)	NOT NULL	Numéro de fax (format international) d'un acteur ou contributeur.
nom_organisation	character varying	NOT NULL	nom de l'entreprise ou l'organisation à laquelle le contributeur appartient.
position_individu	character varying(255)	NOT NULL	Position du contributeur au sein de cette entreprise ou organisation
<b>login</b>	<b>character varying(20)</b>	<b>PK</b>	<b>identifiant system unique du contributeur</b>
passwd	character varying(16)	NOT NULL	mot de passe du contributeur.

Tableau 51. La table RESPONSABLE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idcitation	integer	UNIQUE	identifiant system unique de ce document, cette oeuvre
titre	integer		titre du document, oeuvre, source - obligatoire
date_citation	integer	FK vers DATE_HISTORIQUE	date de validité du document (date historique) - obligatoire
auteurs	timestamp	NOT NULL	prenom et nom des auteurs, séparés par un point-virgule (;)
edition	boolean	NOT NULL	nom de l'édition
edition_date	boolean	NOT NULL	date (date système) de l'édition - précision au jour
format_oeuvre	character varying	NOT NULL	format sous lequel le document est disponible : papier, électronique, CD, etc.
nom_serie	character varying(255)		nom dans la série
numero_serie	character varying(255)		numéro dans la série
page_serie	character varying(255)		page dans la série
autres_details	character varying(255)		autres détails liés à ce document
titre_collection	character varying(255)		titre de la collection dans laquelle on trouve le document
isbn	character varying(255)		numéro ISBN de la source (si livre)
issn	character varying(255)		numéro ISSN de la source (si revue, périodique, journal, collection)

Tableau 52. La table CITATION.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idprovenance	integer	PK	identifiant système unique
idcontributeur	integer	FK vers RESPONSABLE	Contributeur
idtype_lacune	integer	FK vers TYPE_LACUNE	Définition en français
date_maj	timestamp	NOT NULL	date de la mise à jour (date de transaction, DateTime format)
hypothese	boolean	NOT NULL	vrai si le contributeur émet une hypothèse (information à confirmer)
lacune	boolean	NOT NULL	vrai si l'information est manquante
commentaire	character varying	NOT NULL	commentaire libre

Tableau 53. La table PROVENANCE\_INFORMATION.

Comme une provenance peut être associée à plusieurs citations, et que plusieurs citations s'appliquent à plusieurs provenance, une table de jointure PROVENANCE\_CITATION (n : m) est créée.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idprovenance	integer	FK vers PROVENANCE_INFORMATION	Provenance
idtype_lacune	integer	FK vers CITATION	Œuvre citée

Tableau 54. La table PROVENANCE\_CITATION.

#### 4.1.2. Les tables de nomenclature

Ce schéma modélise l'aspect hiérarchique de l'organisation territoriale, et versionne les différents attributs d'une entité.

##### 4.1.2.1. Tables PERIODE, DIVISION, SUBDIVISION, ZONAGE

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
code	character varying(1)	PK	
description	character varying(255)		
nom	character varying(255)		
date_debut	integer	FK vers DATE_HISTORIQUE	Début de période
date_fin	integer	FK vers DATE_HISTORIQUE	Fin de période

Tableau 55. La table PERIODE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
code	character varying(1)	PK	
description	character varying(255)		
nom	character varying(255)		

Tableau 56. La table DIVISION.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idnature_subdivision	serial	PK	Id
idperiode	integer	FK sur PERIODE	Periode
code	character varying(1)	NOT NULL	
description	character varying(255)		
nom	character varying(255)		
date_debut	integer	FK vers DATE_HISTORIQUE	Début de subdivision
date_debut_exacte	integer	FK vers DATE_HISTORIQUE	Début exact de subdivision
date_fin	integer	FK vers DATE_HISTORIQUE	Fin de subdivision
date_fin_exacte	integer	FK vers DATE_HISTORIQUE	Fin exacte de subdivision

Tableau 57. La table SUBDIVISION.

Le triplet (idnature\_subdivision, idperiode, code) doit être unique dans la table **SUBDIVISION**. Ceci pour garantir l'unicité du code 1 du code historique. Egalement, on vérifie également que les dates de début et fin de subdivision soient incluses dans l'intervalle temporel défini par la période que référence cette subdivision.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
<b>idzonage</b>	serial	<b>PK</b>	<b>Id</b>
idsubdivision	integer	FK sur SUBDIVISION	subdivision
code	integer	NOT NULL	identifiant unique du zonage dans la subdivision considérée – valeur du code2 du code historique.
nom	character varying(255)		nom du zonage dans la subdivision considérée (deux zonages de même noms, mais de niveau différents peuvent co-exister dans une subdivision)
niveau	integer	NOT NULL	Niveau dans la subdivision. niveau dans la subdivision : 1 pour le découpage le moins fin, et croit lorsque le niveau correspond à un découpage plus fin qui s'emboite dans un zonage de plus haut niveau - Le niveau peut-être mise à jour. Si le zonage ne s'inscrit pas dans une hiérarchie, son niveau est -1
est_niveau_base	boolean	NOT NULL	vrai seulement pour le niveau plancher (qui forme le découpage le plus fin de cette subdivision).
metadata	integer	FK sur PROVENANCE_INFORMATION	Métadonnées relative à ce zonage (dans quelles sources bibliographiques est-il fait mention de ce découpage ? ...)

Tableau 58. La table ZONAGE.

Le couple (idsubdivision, code) est unique pour garantir l'unicité du code du zonage dans une nomenclature.

**4.1.2.2. Tables ENTITE\_TERRITORIALE, CODE\_HISTORIQUE, APPARTENANCE, ZONAGE\_COMPOSITION et RECENSEMENT**

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
<b>identite_territoriale</b>	<b>serial</b>	<b>PK</b>	<b>Id</b>
Code_INSEE_1999	character varying(50)	Peut-être NULL si l'entité n'existe pas encore en 1999 ou si l'entité ne fait pas partie de la nomenclature INSEE.	Code INSEE long de l'entité en 1999 si elle existait en 1999. Si l'entité a disparue, il est valué par le code court INSEE 1999 auquel le chef-lieu de l'entité a été rattaché.
Num_ordre	Integer	<b>UNIQUE.</b> Peut-être null si l'entité n'existe pas encore en 1999 ou si l'entité ne fait pas partie de la nomenclature INSEE.	Identifiant de l'entité dans la base de cassini. Aide les mises à jour et les contrôles de cohérence avec la base de Cassini.
Debut_validité	integer	FK vers DATE_HISTORIQUE	Début de validité - première date de rencontre dans les archives
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin d'activité connue (date d'absorption, ou aujourd'hui si encore active)
idmetadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations sur l'unité

Tableau 59. La table ENTITE\_TERRITORIALE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
Id_entite		FK sur ENTITE_TERRITORIALE	Identifiant de l'entité territoriale
Code1	character varying(3)	NOT NULL	Concaténation des codes de Periode, division, sous-division (sur 3 digits)
Code2	Integer	NOT NULL	Code identifiant le zonage (sur 2 digits)
Code3	Integer	NOT NULL	numero de l'entité de référence (sur 3 digits)
Code4	Integer	NOT NULL	Index de l'unité dans son groupe de référence. (sur 4 digits)
Code5	Integer	NOT NULL	motif de transformation (sur deux digits)
Code6	Integer	NOT NULL	index de transformation (numero de version), sur 2 digits
Code7	Integer	NOT NULL	niveau du zonage
codecomplet	character varying(30)	NOT NULL	Calculé, c'est une chaîne alphanumérique qui correspond au code historique de l'unité.
Debut_validité	integer	FK DATE_HISTORIQUE vers	date de début de validité du code historique (initié à la date de création de l'entité dans les archives, puis à chaque nouvelle transformation).
Fin_validité	integer	FK DATE_HISTORIQUE vers	date de fin de validité du code historique (quand l'entité se transforme de nouveau).

Tableau 60. La table CODE\_HISTORIQUE.

La clé primaire du CODE\_HISTORIQUE est (codecomplet, début\_validité, fin\_validité).

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
inf_entite	integer	FK sur ENTITE_TERRITORIALE	Entité territoriale
sup_entite	integer	FK sur ENTITE_TERRITORIALE	Entité supérieure d'un autre niveau
inf_zonage	Integer	FK sur NIVEAU	Identifiant du niveau de l'entité
sup_zonage	Integer	FK sur NIVEAU	Identifiant du niveau de l'entité supérieure
sup_referente	boolean		VRAI si l'entité sert à numérotter ses entités inférieures (cas d'un département par exemple).
index_entiteinf	integer		Numéro de l'entité inférieure dans l'ensemble détenu par l'entité supérieure
Début_validité	integer	FK vers DATE_HISTORIQUE	Début de validité
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin de validité
idmetadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 61. La table APPARTENANCE.

La clé primaire de APPARTENANCE est (inf\_entite, sup\_entite, inf\_zonage, sup\_zonage, début\_validité, fin\_validité).

Pour calculer le code 6 d'une entité de niveau x, il faut lire la table APPARTENANCE pour trouver son entité supérieure y. Ensuite, il faut compter le nombre d'entités de niveau x qui ont cette entité supérieure y durant la même période, et incrémenter d'une unité ce nombre.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
id_entite	integer	FK	Entité territoriale
idzonage	integer	FK sur ZONAGE	Zonage d'appartenance
Début_validité	integer	FK vers DATE_HISTORIQUE	Début de validité
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin de validité
idmetadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 62. La table ZONAGECOMPOSITION.

La clé primaire de ZONAGECOMPOSITION est (id\_entite, idzonage, début\_validité, fin\_validité).

Les **recensements** sont conservés dans une table dédiée. Si le recensement n'a pas pu se faire pour l'entité, l'entrée **lacune\_type** renvoie vers la table TYPE\_LACUNE\_RECENSEMENT pour en connaître la la raison. Le recensement n'est liée qu'à une entité (un nom, un territoire, un chef-lieu, une entité supérieure et un niveau).

Par ailleurs, les dates de recensement ne sont pas homogènes sur tout le territoire, suivant les époques. Ainsi, selon le site Web : « Pour l'Alsace-Lorraine, les recensements de la période 1870-1919 ont eu lieu aux années 0 et 5 sauf 1871, et 1915 qui n'a pas été réalisé. Pour Nice et la Savoie, les recensements de la période 1814-1860 ont eu lieu en 1815, 1822, 1838, 1848 et 1858. ». Cette information est réutilisable pour donner les vraies dates de recensement. Cependant, dans l'optique de construire des séries démographiques complètes, un numéro d'ordre est conféré aux recensements pour comparer des recensements ayant même numéro d'ordre. Ainsi, le recensement de 1870 de Strasbourg coinciderait avec celui de 1870 de Paris.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idrecensement	serial	PK	Id
idEntitéTerritoriale	Integer	FK vers IDENTITE_TERRITORIALE	Identifiant de l'entité territoriale qui a été recensée.
ordre_recensement	Integer		Ordre du recensement
date	integer	FK vers DATE_HISTORIQUE	Date du recensement
valeur	integer		Nombre d'habitants ou NULL si inconnue
lacune_type	integer	FK vers TYPE_LACUNE_RECENSEMENT	Explique la raison de la valeur NULL pour le recensement.

Tableau 63. La table RECENSEMENT.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
Idtype_lacune_recensement	serial	PK	Id
valeur	character varying		
nom	character varying		
description	character varying		

Tableau 64. La table TYPE\_LACUNE\_RECENSEMENT.

#### 4.1.2.3. Tables TOPONYME, et TYPE\_TOPONYME

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
Id_toponyme	serial	PK	Id
id_type	integer	FK sur TYPE_TOPONYME	Type de toponyme
id_cheflieu	integer		Identifiant du chef-lieu qui porte ce toponyme. Si null, c'est le nom d'entité.
Id_entite	integer		Entité qui porte ce toponyme (son numéro d'ordre). Si null, c'est le toponyme d'un chef-lieu
nom	character varying(255)	NOT NULL	Nom

Debut_validité	integer	FK vers DATE_HISTORIQUE	Début de validité
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin de validité
metadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 65. La table TOPONYME.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
code	character varying(1)	PK	Identifiant du type
nom	character varying		Nom du type
description	character varying		Description complète

Tableau 66. La table TYPE\_TOPONYME.

#### 4.1.2.4. Tables CHEFLIEU et GOUVERNANCE

La table CHEFLIEU (Tableau 67) recense tous les chefs-lieux, décrits par une description éventuellement, un champ de type booléen indiquant si le chef-lieu « x » est une entité territoriale, un champ indiquant s'il est aussi de nature ponctuelle. Si le chef-lieu « x » est une entité territoriale, l'identifiant de l'identité territoriale chef-lieu est indiqué dans la table ENTITE\_EST\_CHEFLIEU et référence une entité de la table ENTITE\_TERRITORIALE. Si le chef-lieu « x » est un point, sa géométrie doit être recherchée dans la table CENTRE. Si « x » est une entité territoriale, il suffit de lire sa géométrie dans la table LIMITE consacrée aux géométries d'entités territoriales (voir la section 4.1.3 page 60). Dans les deux cas, un point calculé peut-être conservé. Si la longitude et la latitude du chef-lieu sont connues, elles sont renseignées dans le CENTRE, quelle que soit la nature du chef-lieu.

Si l'entité n'existe pas car c'est un élément sous le niveau de base qui n'a pas encore été créé, elle est créée dans la table entité, avec pour nom le nom de ce chef-lieu, (type Usage), et pour entité supérieure l'entité dont elle est chef-lieu. Par exemple, un hameau dont on ne trouve trace que comme chef-lieu (ancien ou nouveau) d'une commune dans la table TCL\_COMMUNE, est créé comme entité infra\_communale, et appartient à la commune dont il est le chef\_lieu. Il n'a pas de territoire connu, son niveau est infra\_communal, et son nom est celui décrit dans TCL\_COMMUNE. Ancien\_CL ou TCL\_COMMUNE.nouveau\_CL, et le type de ce toponyme est « Usage ».

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
idchef_lieu	serial	PK	Id d'un chef-lieu,
statut	character varying(255)		Nom générique du centre de gouvernance : Capitale, Sous-Préfecture, Préfecture, chef-lieu de canton, etc.
description	character varying(255)		commentaire
est_un_point	Boolean	NOT NULL	Vrai si le chef-lieu est un point, faux si c'est une entité territoriale chef-lieu d'autres entités.
est_une_entite	Boolean	NOT NULL	Vrai si le chef-lieu est une entité territoriale chef-lieu d'autres entités
Début_validité	integer	FK vers DATE_HISTORIQUE	Début de validité
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin de validité
metadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 67. La table CHEFLIEU.

La table ENTITE\_EST\_CHEFLIEU enregistre le fait qu'une entité soit cheflieu, avec des contraintes sur les clés étrangères. Par exemple, Grenoble est une préfecture qui est chef-lieu (Tableau 68).

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
Id_chef_lieu	integer	FK sur CHEF_LIEU	Id d'un chef-lieu,
Id_entite	integer	FK sur IDENTITE_TERRITORIALE	Id de l'entité territoriale gouvernée. Voir l'analyse § 3.3.1 page 46

Tableau 68. La table ENTITE\_EST\_CHEFLIEU.

La table GOUVERNANCE (Tableau 69) enregistre le lien « est chef-lieu de » entre les chefs-lieux et les entités territoriales gouvernées, avec des contraintes sur les clés étrangères. Par exemple, Grenoble est une préfecture qui est chef-lieu de toutes les communes du département de l'Isère.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
Id_chef_lieu	integer	FK sur CHEF_LIEU	Id d'un chef-lieu,
Id_entite	integer	FK sur IDENTITE_TERRITORIALE	Id de l'entité territoriale gouvernée. Voir l'analyse § 3.3.1 page 46
Début_validité	integer	FK vers DATE_HISTORIQUE	Début de validité de la relation de gouvernance
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin de validité de la relation de gouvernance
metadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 69. La table GOUVERNANCE.

#### 4.1.2.5. Tables EVENEMENT, EVENEMENT\_ENTITE\_MOTIF et MUTATION\_MOTIF

La table EVENEMENT (Tableau 70) sert à retracer les événements qui provoquent l'apparition ou la disparition d'une entité, un changement de nom, d'appartenance, ou de statut (transfert de chef-lieu).

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
id_evenement	serial	PK	Identifiant de l'événement qui peut indexer plusieurs transformations à la fois à la même date
index_evenement_cassini	integer	FK sur MUTATION_MOTIF	Motif de l'événement
date_mutation	integer	FK vers DATE_HISTORIQUE	Date de l'événement
commentaire	character varying(255)	NON NULL	Description à l'intention des utilisateurs
metadata	integer	FK sur PROVENANCE_INF ORMATION	Source des informations

Tableau 70. La table EVENEMENT.

La table EVENEMENT\_ENTITE\_MOTIF (Tableau 71) permet de retranscrire ce qui arrive à chaque entité lors d'un même événement territorial. Le motif, compatible INSEE, décrit précisément la transformation, dans la table MUTATION\_MOTIF (Tableau 72). C'est le code 5 du code historique de la version de l'entité territoriale qui débute avec cet événement. MOTIF\_GENERIQUE, renseigné avec un code variant entre 1 et 9, décrit l'événement de façon générique pour les entités impliquées dans un même événement. Par exemple, lors de la création d'une commune (motif\_générique=2), certaines communes cèdent des parcelles (motif =22) tant dis qu'une commune au moins est créée (motif =20). La clé primaire de EVENEMENT\_ENTITE\_MOTIF est (id\_evenement, id\_entite, motif).

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
id_evenement	serial		Identifiant de l'événement qui peut indexer plusieurs transformations à la fois à la même date
id_entite	integer	FK sur IDENTITE_TERRITORIALE	identifiant unique de l'entité
motif	integer	FK sur MUTATION_MOTIF	Motif qui s'applique à l'entité impliquée dans l'événement
motif_generique	integer	FK vers DATE_HISTORIQUE	Motif générique qui s'applique à l'événement
metadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 71. La table EVENEMENT\_ENTITE\_MOTIF.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
code	integer		
nom	character varying(255)		
definition	character varying(255)		

Tableau 72. La table MUTATION\_MOTIF.

#### 4.1.3. Les tables géographiques

Nous proposons deux tables géographiques :

- une table LIMITÉ qui stocke les contours des entités surfaciques (des MULTI-POLYGONES), avec leur projection,
- une table CENTRE qui stocke les coordonnées des centres (POINT), avec leur projection.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
<b>Id_centre</b>	serial	<b>UNIQUE</b>	<b>Id du centre</b>
<b>Id_projection</b>	integer	FK sur SPATIAL_REF_SYS	Id de la projection utilisée
<b>id_cheflieu</b>	integer	<b>PK, FK sur CHEFLIEU</b>	<b>Id du chef-lieu</b>
geometrie	Geometrie	FK IDENTITE_TERRITORIALE sur	Coordonnées du centre : ville ou centre administratif
longitude	character varying(255)		longitude en degré/minutes/secondes du chef-lieu si connue
latitude	character varying(255)		latitude en degré/minutes/secondes du chef-lieu si connue
metadata	Integer	FK PROVENANCE_INFORMATION sur	Source des informations

Tableau 73. La table CENTRE.

CHAMP	TYPE	CONTRAINTE	SIGNIFICATION
<b>Id_limite</b>	serial	unique	Id de la géométrie
<b>Id_projection</b>	integer	FK sur PROJECTION	Id de la projection utilisée
contour	Geometry	NOT NULL	Multi-polygone définissant le territoire de l'unité.
centroid	Geometry		Coordonnées calculées du centre du polygone (par la fonction <i>centroid</i> de postgis)
Surface_legale	double		Surface administrative exprimée en hectares, doit être positive si elle est connue.
Altitude_min	double		Altitude du point le plus bas de l'entité, en mètres
Altitude_max	double		Altitude du point le plus élevé de l'entité, en mètres
Surface_calcule	integer		Calculée à partir des géométries, exprimée en km <sup>2</sup> , elle peut différer de la surface légale et dépend du niveau de généralisation des géométries.
<b>Id_entite</b>	integer	FK ENTITE_TERRITORIALE sur	Id de l'entité territoriale qui a cette géométrie
Debut_validité	integer	FK vers DATE_HISTORIQUE	Début de validité de cette limite.
Fin_validité	integer	FK vers DATE_HISTORIQUE	Fin de validité de cette limite.
metadata	integer	FK sur PROVENANCE_INFORMATION	Source des informations

Tableau 74. La table LIMITÉ.

La clé primaire de LIMITÉ est (id\_entite, debut\_validité, fin\_validité)

## 4.2. Migration de Cassini vers GeoPeople

Nous détaillons la procédure de remplissage de GeoPeople : contraintes sur l'ordre des insertions, modalités de calcul de certains champs. Toute la base GeoPeople est encodée en **UTF8**.

L'ordre de remplissage des différentes tables et champs de GéoNomenclature est spécifié ici, et les algorithmes de calcul de certains champs sont détaillés lorsque la correspondance entre les données n'est pas évident.

### 4.2.1. Correspondance des types

Quelques incompatibilités de type entre la base MySQL dont est issue la base Cassini et la base GeoNomenclature sont notées et documentées dans le Tableau 75.

Par exemple, les types int(10) sont des entiers (*integer*), sous postgres.

Il faut éviter de spécifier des longueurs de mantisse sur les entiers ou sur les réels, dans un souci de portabilité entre les différents SGBD et systèmes d'exploitation.

Type MySQL	Type Postgres	Commentaire ou documentation
Int(10)	Integer	<a href="http://sql.toutestfacile.com/cours/les_types_sql_1.php5">http://sql.toutestfacile.com/cours/les_types_sql_1.php5</a>
tinyint(1)	Integer	<a href="http://dev.mysql.com/doc/refman/5.0/en/numeric-types.html">http://dev.mysql.com/doc/refman/5.0/en/numeric-types.html</a>
smallint(5)	Integer ou smallint	<a href="http://www.postgresql.org/docs/7.4/static/datatype.html#DATATYPE-INT">http://www.postgresql.org/docs/7.4/static/datatype.html#DATATYPE-INT</a>
Double(15,5)	Float(15) ou real	“For maximum portability, code requiring storage of approximate numeric data values should use <b>FLOAT</b> or <b>DOUBLE PRECISION</b> with no specification of precision or number of digits.” <a href="http://www.postgresql.org/docs/7.4/static/datatype.html#DATATYPE-FLOAT">http://www.postgresql.org/docs/7.4/static/datatype.html#DATATYPE-FLOAT</a>
Varchar(n), avec n entier	character varying(n)	<a href="http://dev.mysql.com/doc/refman/5.0/en/char.html">http://dev.mysql.com/doc/refman/5.0/en/char.html</a> chaîne de caractère limitée à n caractères
longtext	character varying or text	<a href="http://dev.mysql.com/doc/refman/5.0/en/blob.html">http://dev.mysql.com/doc/refman/5.0/en/blob.html</a> “Une colonne <b>LONGTEXT</b> ou <b>LONGBLOB</b> peut contenir au maximum 4294967295 ou 4 Go ( $2^{32} - 1$ ) caractères. Jusqu'en version 3.23 le protocole client/serveur et les tables MyISAM avait une limite de 16 Mo par paquet de communication pour une ligne de table. Depuis les versions 4.x, la taille maximale d'un <b>LONGTEXT</b> ou <b>LONGBLOB</b> dépend de la taille maximal de paquet de communication pour le protocole de communication, et de la mémoire disponible. » <a href="http://www.postgresql.org/docs/7.4/static/datatype-character.html">http://www.postgresql.org/docs/7.4/static/datatype-character.html</a> « If you desire to store long strings with no specific upper limit, use <b>text</b> or <b>character varying</b> without a length specifier, rather than making up an arbitrary length limit.”

Tableau 75. Traduction des types entre MySQL et Postgres.

### 4.2.2. Analyse des toponymes

Ce paragraphe détaille comment les types de toponyme sont déterminés à partir de la base Cassini.

Les toponymes dans Cassini :

- toponyme.nom\_toponyme : une version du nom, officielle, révolutionnaire, ou autre (locale ou variante).
- commune.nom\_ancien : nom ancien (avant 1793)

Déterminer le type :

- si toponyme.nom\_toponyme est comme commune.nom\_revo : nom révolutionnaire
- si toponyme.info\_cgt\_topo vaut "ACTE" : version officielle

Déterminer la(s) source(s) :

- par défaut, la source de toutes les informations est aussi le BD EHESS Cassini
- si commune.nom\_an3 est comme toponyme.nom\_toponyme : vu dans le recensement 1793 (1793)
- si toponyme.topo\_1794 vaut E : vu dans les actes an II (1793)
- si toponyme.topo\_1801 vaut E : vu dans le Bulletin des lois 1801 (1801)
- si toponyme.topo\_1999 vaut E : vu dans la BD INSEE 1999 (1999)

Déterminer la période de validité :

Si le nom est ancien, début\_période vaut (avant 1793) et fin\_période vaut (1793).

Si c'est une variante révolutionnaire, elle est datée entre 1793 et 1801.

Sinon:

- Début\_période : entre toponyme.annee\_debut\_possession\_topo et toponyme.annee\_fin\_cgt\_topo
- Fin\_période : toponyme.annee\_fin\_possession\_topo

Nous avons considéré que le nom était 'officiel' si et seulement si il y avait la mention ACTE sur son entrée dans toponymes, et que les colonnes topo\_x renseignaient sur les sources de l'information. Contrairement à Cassini, le fait que le nom apparaisse dans la BD INSEE 99 n'en fait pas un toponyme de type officiel. Ainsi, si cette variante n'a pas été trouvée dans un acte de changement de nom, il reste une variante.

### Quelques exemples :

#### • **Eaux-Bonnes**

ancien(s)	information non disponible actuellement
nom(s) rév.	information non disponible actuellement
an II	1793, Aas
Bull. des Lois	1801, Aast Aas 1861, Eaux-Bonnes

Dans GeoPeuple :

- Aas était le toponyme officiel entre 1793 et 1861, suivant les sources « Acte an II », la base cassini
- Aast était une variante remarquée dans le Bulletin des lois 1801
- Eaux-Bonnes est le toponyme officiel depuis 1861, et apparaît dans les sources Insee 1999 et la Base de Cassini.

#### • **Abancourt**

ancien(s)	information non disponible actuellement
nom(s) rév.	Abancourt-la-Montagne
an II	1793, Abbancourt
Bull. des Lois	1801, 1823, Abancourt

Dans GeoPeuple

- Abancourt est le toponyme officiel depuis 1823, présent dans la base cassini et la base INSEE 1999

- Abbancourt était le toponyme officiel entre 1793 et 1823, suivant les sources « Acte an II », la base cassini
- Abancourt-la-Montagne est une variante révolutionnaire entre 1793 et 1800.

- **Aboën**

ancien(s)	information non disponible actuellement
nom(s) rév.	information non disponible actuellement
an II	1793, Aboin et Salineaux
Bull. des Lois	1801, ... 1872, Aböen Aboën

Dans GéoPeuple,

- Aböen est le toponyme officiel depuis 1872, présent dans la base cassini
- Aboën est une variante depuis 1872, présent dans la base cassini et la base INSEE 1999
- Aboin et Salineaux était le toponyme officiel entre 1793 et 1872, suivant les sources « Acte an II », la base cassini

- **Saint-Jean-de-Niost**

ancien(s)	Gourdans
nom(s) rév.	Niost
an II	1793, Saint Jean de Niord
Bull. des Lois	1801, Saint-Jean-de-Niort Saint-Jean-de-Niost

Dans GéoPeuple,

- Saint Jean de Niord est le toponyme officiel depuis 1793, présent dans la base cassini et les sources « Acte an II ».
- Saint-Jean-de-Niost est une variante, présent dans la base cassini et la base INSEE 1999
- Saint-Jean-de-Niort est une variante, présent dans la base cassini et et les sources « Bulletin des lois 1801 ».
- Niost est une variante révolutionnaire entre 1793 et 1801
- Gourdans est un toponyme d'avant 1789

- **Vensat**

ancien(s)	information non disponible actuellement
nom(s) rév.	information non disponible actuellement
an II	1793, Saint Jean de Venssat 1800, Vensat
Bull. des Lois	1801, ...

Dans GeoPeuple,

- Vensat est le toponyme officiel depuis 1800, présent dans la base cassini et la base INSEE 1999.
- Saint-Jean-de-Vensat est une variante, présente dans la base cassini
- Veussat est une variante, présente dans la base cassini et les sources « Bulletin des lois 1801 ».
- Saint Jean de Venssat était le toponyme officiel entre 1793 et 1800, d'après l'Acte an II et la base Cassini

### 4.2.3. Migration des événements

Tous les événements de la table EVOLUTION\_TERRITORIALE sont importés en premier dans la table EVENEMENT. Le commentaire est non null seulement pour les cas de transfert de chef-lieux entre communes (le champ NATURE\_TCL de la table TCL\_COMMUNE dans Cassini\_EHESS est copié) ou de changement de nom.

Événements déduits par un programme car non conservés dans la base de Cassini :

- Le rétablissement : c'est une création qui vient après une fusion.
- Le changement de nom : les événements de changement de nom (Officiels) ne sont pas présents dans Cassini. Il faut les calculer et vérifier s'ils ont eu lieu au cours d'une fusion, d'un rétablissement, d'un transfert de chef-lieu.<sup>14</sup>
- Les changements de nom de chef-lieu : extraits à partir du traitement du champ STATUT\_ADMIN de la table COMMUNE.STATUT\_ADMIN.

### 4.2.4. Crédation des limites des communes

Les géométries anciennes des communes ont été produites par Hervé LeBras à partir de la BDCarto 1999. Il exporte ses fonds hypothétiques dans un format spécial.

Nous nous sommes basés sur les limites BDCarto 2011 projetées en Lambert93, avec des attributs encodés en UTF8, pour l'import des géométries actuelles en base.

#### 4.2.4.1. Crédation des limites actuelles à partir de la BDCarto 2011

Attributs de Communes.Shp de la BDCarto

INSEE\_COMM : 69108

NOM\_COMM : en capitales, sans accents

Superficie : donnée en ha

Population : en 2011, totale (peut-être).

La correspondance se fait par l'attribut code\_INSEE du shapefile et le code\_insee\_court de la BD Cassini.

Exemple :

```
INSEE_COMM = 69108
code_insee_court = 69108
code_insee_1999 = 69 2 05 108
```

Requête :

```
select c.num_ordre, c.label_commune, c.surface, c.a_min, c.a_max,
the_geom
from commune c, limitescommunes_bdtopo_193 lim
where lim.insee_comm=c.code_insee_court and c.surface is not null
```

Première procédure :

- Convertir les shapefiles en UTF8 (vis QGIS).  
On obtient Geometries\_communesBDTOPO.shp
- Créer par import du premier shapefile la table limitesCommunes\_BDTOPOL93

<sup>14</sup> Attention, si deux sortes d'événements se produisent à la même date, et qu'ils donnent lieu à un changement de nom, l'événement « changement de nom » est enregistré en double.

```

/usr/local/pgsql-9.0/bin/shp2pgsql -s 2154 -c
Geometries_communesBDTOPO.shp limitesCommunes_BDTOPO_L93 | psql -U
postgres -d Cassini_EHESS
    - Importer les suivantes avec append (a) dans cette même table
/usr/local/pgsql-9.0/bin/shp2pgsql -W LATIN1 -s 2154 -a BDC_3-
0_SHP_LAMB93_21-ED103/ADMINISTRATIF/COMMUNE.shp
limitesCommunes_BDTOPO_L93 | psql -U postgres -d Cassini_EHESS

```

Attention, la colonne "statut" de la table limitesCommunes\_BDTOPO\_L93 doit être étendue (25 caractères). Les scripts test\_\*.sql doivent être édités :

- SET CLIENT\_ENCODING TO LATIN1;

#### Seconde procédure :

Créer les scripts d'insertion test\_\*.sql

- /usr/local/pgsql-9.0/bin/shp2pgsql -W LATIN1 -s 2154 -a BDC\_3-
0\_SHP\_LAMB93\_21-ED103/ADMINISTRATIF/COMMUNE.shp
limitesCommunes\_BDTOPO\_L93 > test\_XYZ.sql

Exécuter les script test\_\*.sql après avoir exécuté la commande suivante :

```

CREATE TABLE "limitescommunes_bdtopo_193" (gid serial PRIMARY KEY,
"id_bdcarto" numeric(10,0),
"nom_comm" varchar(50),
"insee_comm" varchar(5),
"statut" varchar(25),
"x_commune" int4,
"y_commune" int4,
"superficie" numeric(10,0),
"population" int4,
"insee_cant" varchar(2),
"insee_arr" varchar(1),
"nom_dept" varchar(30),
"insee_dept" varchar(2),
"nom_region" varchar(30),
"insee_reg" varchar(2));
SELECT
AddGeometryColumn('','limitescommunes_bdtopo_193','the_geom','2154',
'MULTIPOLYGON',2);

```

Avec un système d'import par jointure sur le code code\_insee\_court, une géométrie (la géométrie actuelle) est importée pour toutes les entités présentes et toujours active dans Cassini. La fonction à appeler est :

```
fillLimites(_dbConnectionCassini,_dbConnectionGeoPeople);
```

Si certaines entités ont disparues ou ont été différentes dans le passé, elles n'ont pas encore de géométrie.

Ainsi, on récupère 17861 limites d'entités. Pour quelques entités, ces limites sont en doubles dans la BDCarto (40229, 20342, 12893, 36092, 15053) mais elles sont bien récupérées dans GeoPeople. Plus précisément, elles sont en double dans la table LIMITESCOMMUNES\_BDTOPO\_L93 parce que les shapefiles utilisés pour importer les données ont quelques recouvrements.

Cependant, certains numéros courts utilisés dans la BDCarto n'apparaissent pas dans la base de données Cassini\_EHESS. L'annexe 5.2 page 72 détaille et analyse ces cas.

#### **4.2.4.1. Cration des limites anciennes a partir des hypothe LeBras**

Les fichiers qui contiennent les gomtries produites par EHESS sont dans un format particulier. Pour chaque commune ancienne, qui a t  absorb e ou a c d  des parcelles pour la cration d'autres communes, un fichier donne l'ent te dcrivant la commune, puis une liste de points qui constituent son contour.

L'ent te contient :

- le nombre de points du polygone,
- le code insee court de la commune (a laquelle elle est rattach e),
- un numero de ligne correspondant a la ligne d'un fichier Excel qui r pertorie toutes les communes trait es par Claude Motte,
- une date ou un couple de dates : la premi re est le d but de validit  de la gomtrie (1791 est parfois indiqu , mais c'est en r alit  1793 qu'il faut prendre), la seconde la date de fin de validit , qui correspond a l'vnement ; si il n'y a qu'une date de pr sente, c'est la date de l'vnement.
- le nom de la commune (a l'poque).

nbrePoints	Code insee	N� ligne	Date(s)	Nom
278	35222	50630	1800-1804	Pleine-Foug�res
402	4160	2044	1846	Le Bourget

```
278 35222 50630 1800-1804 Pleine-Foug res
307013. 2396592.
307053. 2396642.
307147. 2396787.
307177. 2396892.
307187. 2397042.
307202. 2397102.

...
106 80566 39509 1791-1974 Fieffes-Montrelet
592783. 2564980.
592823. 2565055.
592888. 2565156.

...
```

Pour l'import en base, il manque pour chaque commune son numero d'ordre. Pour cela, il faut associer le numero de ligne de l'ent te d'une gomtrie au numero d'ordre correspondant a cette ligne dans le fichier Excel de Claude Motte.

Par ailleurs ne sont donn es que les limites qui suivent 1801. Tous les vnements qui pr c d nt 1801 n'ont pas t  syst matiquement trait s (soit 21444 limites manquantes au maximum). Nous avons t  retrouver au moins 5181 gomtries, sans compter les entit s qui existent toujours, mais dont la forme a chang  au cours du temps.

Les coordonn es x,y par ligne sont donn es avec r p tition pour les points de jonction et pas de r p tition du premier point en dernier.

1. Utiliser le programme ParseLeBrasGeometries.java qui cr e un shapefile "LeBras.shp" correspondant a ces gomtries.
2. Convertir ce shapefile de la projection LambertIIEtendue (EPSG:27572) a la projection Lambert93 (EPSG:2154) avec les outils Ogr2ogr.
3. Extraire de ce shapefile les gomtries qui sont correctes (ne pas s lectionner la diagonale a droite) : LeBras\_OK\_L93.shp

4. Importer en base le shapefile ainsi modifié : "/usr/local/pgsql-9.0/bin/shp2pgsql -s 2154 -c LeBras\_OK\_L93.shp limitesAnciennesCommunes > importLimitesAnciennesCommunes.sql"
5. Utiliser le programme *fillOldLimites*

#### 4.2.4.2. Mise à jour des dates de validité des limites de commune

Les géométries ont pour période [1793, maintenant] par défaut.

- Si la commune n'existe que depuis la date X, (événement de création), alors on met à jour la date de début de validité : [X, maintenant].
- Si la commune a été absorbée à la date X, (événement d'absorption), alors on met à jour la date de fin de validité : [1793, X]. Et le champs 'the\_geom' prend NULL.
- Si la commune a échangé des parcelles avec d'autres communes, la date de validité des géométries actuelle est mise à jour en fonction de la dernière date de modification de limites.
- Dans les cas de rétablissement à la date Y : les anciennes limites sont reprises, et valides jusqu'au prochain événement.

La lecture des événements permet de mettre ensuite les dates de début des limites insérées à jour. Cet algorithme est mis en œuvre par la procédure suivante :

`migration.checkLimites(dbConnectionGeoPeople)`

Il s'agit aussi de compléter les géométries manquantes, car elles sont celles d'unités qui ont été divisées, ou bien rétablies. Le schéma Figure 11 décrit les deux cas.

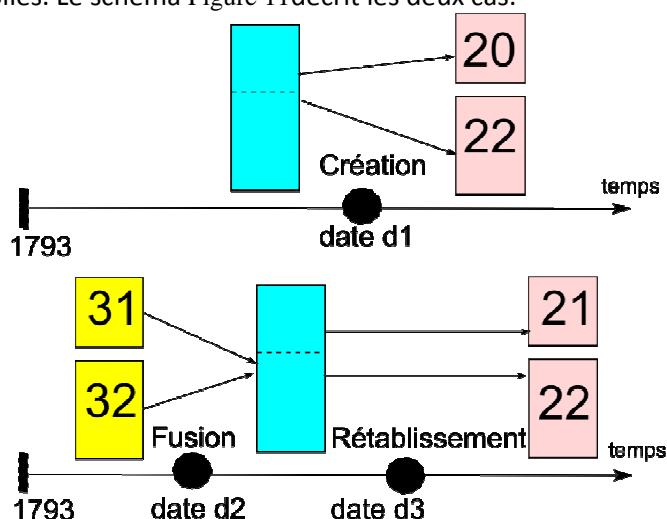


Figure 11. Créations, Absorptions et Rétablissements.

Le premier cas correspond à une création à partir d'une entité (en bleu) qui donne des parcelles (code 22), pour la création d'une nouvelle (en rose, code 21), à la date d1.

Le second cas correspond à une absorption d'une entité (code 31 en jaune) par une autre entité (code 32 en jaune) à la date d2, pour donner une entité plus grande (en bleu), qui a le même identifiant que celle en jaune, code 31. Ensuite, l'entité 31 reprend ses anciennes limites par rétablissement, en rose, et l'entité absorbante cède les parcelles absorbées (code 22 en rose).

L'appel au code `buildOldLimitesBeforeFusion (dbConnectionGeoPeople)` permet de calculer les limites en jaune dans les cas de rétablissement, ainsi que la date de début de validité des entités en bleu.

L'appel au code `buildOldLimitesCreationSimpleRetablissement(dbConnectionGeoPeople)` permet de calculer les limites en bleu dans les deux cas.

Enfin, l'appel à ces fonctions doit être suivi par un appel à `migration.checkLimites(dbConnectionGeoPeople)` qui vérifie et met à jour des dates de limites des entités.

Enfin il est nécessaire de calculer en km<sup>2</sup> la surface de toutes les entités, arrondies à l'entier le plus proche.

```
UPDATE history_limite SET
surface_calcul=round(st_area(contour)/1000000)
```

#### 4.2.5. Déterminer les communes des zones de saisie

Les zones de saisie (sous Cassini puis ensuite l'Etat-Major) sont définies par des shapefiles dessinés à la main qui suivent le contour de ce qui a été saisie à la main dans Cassini et l'extension de la feuille de Cassini.

Ces shapefiles ont été importés en base dans la table **ZONECASSINI** (aussi nommée ultérieurement **METADATA\_ZONES\_SAISIE**) (dont le Tableau 76 liste les attributs non spatiaux).

zone	zone_id
Grenoble-cassini	1
Saint-Malo-cassini	2
Reims-cassini	3
Agen-cassini	4
Grenoble-cassini-complet	11
Saint-Malo-cassini-complet	22
Reims-cassini-complet	33
Agen-cassini-complet	44

Tableau 76. Zones de saisie.

Cette table permet de déterminer par jointure spatiale l'ensemble des communes qui font partie de la zone, et le résultat est sauvegardé dans **zone\_commune\_OK**. On procède par jointure spatiale entre nos zones et les limites actuelles des communes (ce qui sélectionne par exemple Seyssins d'aujourd'hui, num\_ordre 38283) et puis par jointure sur **commune** (table de la base Cassini\_EHESS) pour avoir les communes (actuelles ou anciennes) par jointure sur code insee court. En effet, les communes anciennes absorbées par Seyssins ont comme code court le code insee de Seyssins.

```
CREATE TABLE zone_commune_OK as
(select zone_id, insee_comm, nom_comm, num_ordre, code_insee_court,
c.code_insee_1999, (surface is null) as disparue ,
_st_contains(z.the_geom, lim.the_geom) as inside
from zonescassini z, limitescommunes_bdtopo_193 lim, commune c
where (_st_intersects(z.the_geom, lim.the_geom))
and lim.insee_comm=c.code_insee_court
order by zone_id )
update zone_commune_ok z set id_entite = (select id_entite from
history_entite_territoriale where num_ordre=z.num_ordre)
```

Après import de cette table dans la base GeoPeople, nous la mettons à jour pour lui ajouter l'identifiant de l'entité (colonne id\_entite) dans GeoPeople (plus pratique pour travailler avec).

```
update zone_commune_ok z set id_entite = (select id_entite from history_entite_territoriale where num_ordre=z.num_ordre)
```

#### 4.2.1. Procédure de migration

Il faut d'abord restaurer la base de Cassini, avec PgAdmin, en créant une base de donnée SPATIALE (avec le template\_postgis) encodée en UTF8, comme l'illustre la Figure 12 .

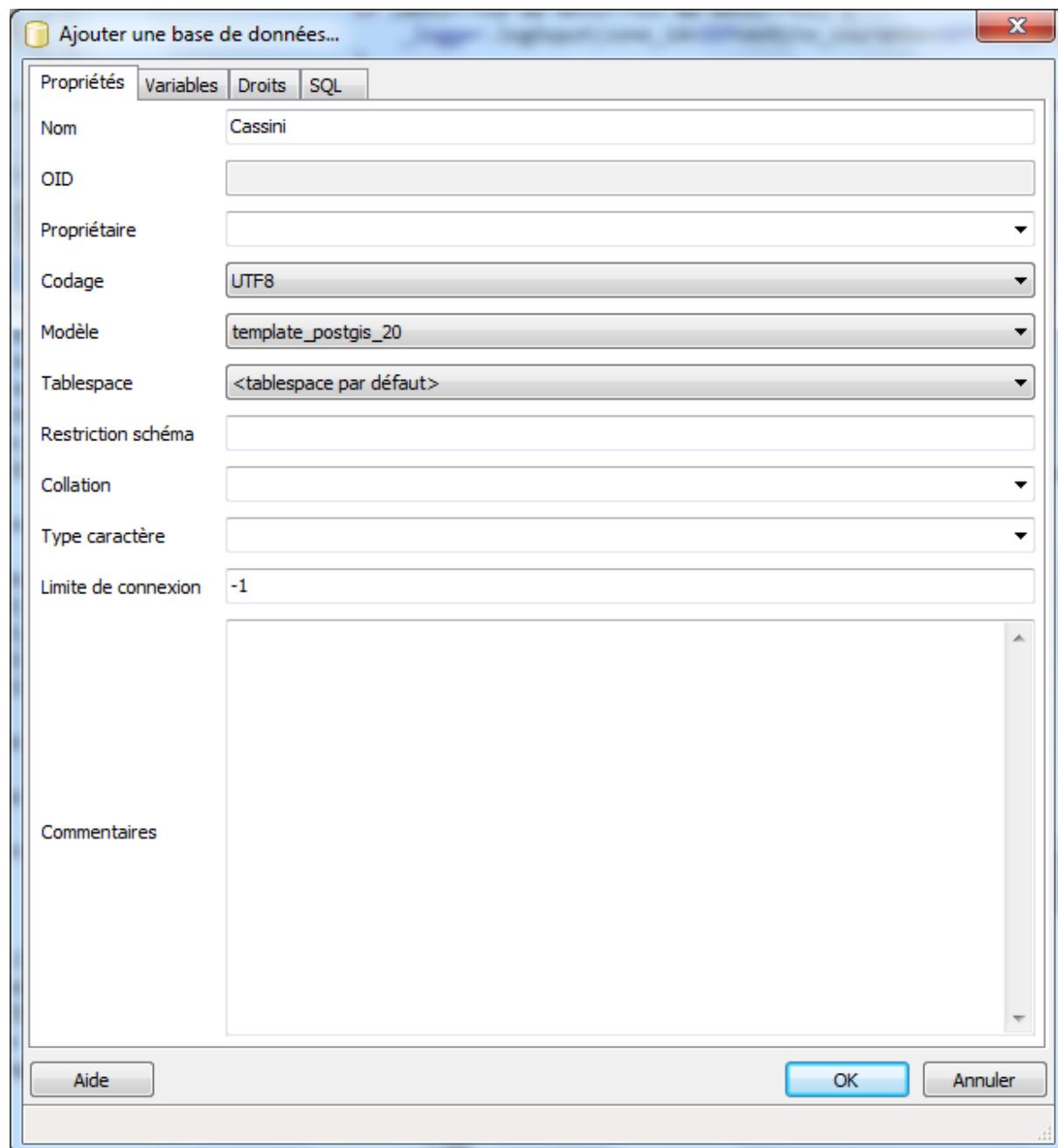


Figure 12. Création d'une base de données spatiale.

La restauration de la première partie (Cassini\_UTF8.backup) importe les 37 tables de la base de données de Cassini originelle. La restauration de la seconde partie (Limite\_UTF8.backup) importe les 4 tables correspondant à l'import de limites anciennes et nouvelles, ainsi qu'à la définition des zones d'étude et des communes qui en font partie.

Ensuite, il faut créer la base de données GeoPeople et sa structure :

- Créer une base de données spatiale vide encodée en UTF8
- Exécuter les scripts suivants :
  - o GeoPeople\_geonoménomination.sql
  - o procéduresGeoPeople.sql

A partir de là, un programme **ImportCassiniEHESS** écrit en Java effectue l'import des données depuis la base Cassini vers la base GeoPeople.

Il crée deux connexions :

- sur la base de Cassini : `_dbConnectionCassini`
- sur la base de GeoPeople : `_dbConnectionGeoPeople`.

Puis :

- importe les communes en tant qu'entités territoriales :  
`fillCommuneAsEntite(_dbConnectionCassini, _dbConnectionGeoPeople)`
- importe les toponymes des communes :  
`fillToponymes(_dbConnectionCassini, _dbConnectionGeoPeople);`
- importe les données de recensement :  
`fillRecensement(_dbConnectionCassini, _dbConnectionGeoPeople) ;`
- importe les limites actuelles des communes :  
`fillLimites(_dbConnectionCassini, _dbConnectionGeoPeople) ;`
- importe les événements associés aux communes : dates et identifiants  
`fillEvent(_dbConnectionCassini, _dbConnectionGeoPeople);`
- importe les événements associés aux communes de type échange de parcelles (code 60 et 62)  
`fillEventParcellesMvnt(_dbConnectionCassini, _dbConnectionGeoPeople);`
- importe les événements associés aux communes de type suppression (code 30, 31 et 32)  
`fillEventSuppression(_dbConnectionCassini, _dbConnectionGeoPeople);`
- importe les événements associés aux communes de type création (code 20, et 22)  
`fillEventCreation(_dbConnectionCassini, _dbConnectionGeoPeople);`
- importe les événements associés aux communes de type transfert de chef-lieu (code 56 et 57)  
`fillEventTransfertChefLieux(_dbConnectionCassini, _dbConnectionGeoPeople);`
- Déduire les événements de changements de noms dûs à des créations, suppressions, ou échanges de parcelles, et les événements de rétablissement (code 21)  
`postProcessingHistoryData(_dbConnectionCassini, _dbConnectionGeoPeople) ;`

Compléter et mettre à jour les dates de limites :

```
checkLimites(dbConnectionGeoPeople) ;
fillOldLimites(_dbConnectionCassini, _dbConnectionGeoPeople) ;
checkLimites(dbConnectionGeoPeople) ;
buildOldLimitesBeforeFusion() ;
checkLimites(dbConnectionGeoPeople) ;
buildOldLimitesCreationSimpleRetablissement(dbConnectionGeoPeople) ;
checkLimites(dbConnectionGeoPeople) ;
```

## 5. Annexes

### 5.1. Liste complète des zonages définis pour GeoNomenclature

idzonage	idsubdivision	code	nom	niveau	est_niveau_base
1	1	1	Région géographique	1	f
2	1	2	pays	1	f
3	1	3	petit pays	1	f
4	1	4	micro-pays	1	f
5	2	1	Communauté urbaine	-1	f
6	2	2	Communauté d'agglomérations	-1	f
7	2	3	Communauté de communes	-1	f
8	2	4	Métropole	-1	f
9	2	5	Pays	-1	f
10	2	6	Parc naturel	-1	f
11	3	1	Région	1	f
12	3	2	Département	2	f
13	3	3	Arrondissement	3	f
14	3	4	Canton	4	f
15	3	90	Commune	9	t
16	3	99	Lieu (élément rattaché à une commune)	99	f
17	4	1	Région	1	f
18	4	2	département d'outre-mer	2	f
19	4	3	Collectivités territoriales	-1	f
20	4	4	Pays constitutifs de la République	-1	f
21	4	5	Dépendances de la République	-1	f
22	5	1	Département (1790)	1	f
23	5	2	district (1790-1795)	2	f
24	5	3	municipalité cantonale (1795-1800)	2	f
25	5	4	canton (1790)	3	f
26	5	90	Commune	9	t
27	7	1	Région apostolique	1	f
28	7	2	Archevêché	2	f
29	7	3	Diocèse	2	f
30	7	4	Doyenné	3	f
31	7	5	nouveau doyenné	3	f
32	7	6	secteur missionnaire	3	f
33	7	7	zone	3	f
34	7	8	unité pastorale	4	f
35	7	9	communauté pastorale	4	f

36	7	90	paroisse	9	t
37	7	91	relais paroissial	9	t
38	7	92	communauté de village	9	t
39	7	99	lieu (élément rattaché à une paroisse)	99	f
40	8	1	Région	1	f
41	8	2	consistoire	2	f
42	8	3	secteur	3	f

Tableau 77. Zonages de GeoNomenclature.

## 5.2. Analyse des cas de limites manquantes

### 5.2.1. Limites actuelles non importées

Requête pour récupérer les cas non importés :

```
(select insee_comm, nom_comm from limitescommunes_bdtopo_193 )
EXCEPT
(select insee_comm, nom_comm from commune, limitescommunes_bdtopo_193 where
code_insee_court=insee_comm )
```

Ce sont les 4 communes suivantes :

```
"35317";"SAINT-SYMPHORIEN"
"51201";"CUISLES"
"14482";"OUEZY"
"31300";"LIEOUX"
```

#### Analyse de SAINT-SYMPHORIEN :

Pour **SAINT-SYMPHORIEN**, il existe pourtant bien une entité avec ce toponyme dans la base de données GeoPeuple, mais son code INSEE est différent (35130). Cette entité n'a pas de limites après l'importation. Elle est sensée avoir disparue d'après la BD Cassini\_EHESS lors d'une fusion en 1973.

```
select * from history_entite_territoriale e, history_toponyme t where
e.id_entite=t.id_entite and t.nom LIKE 'Saint-Symphorien' and
code_insee_1999 LIKE '35%'
select * from history_limite where id_entite=34759
select * from commune where code_insee_court LIKE '35130'
```

#### Analyse de CUISLES :

L'entité avec ce nom ou ce code insee n'existe pas dans la BD Cassini\_EHESS.

```
select * from history_entite_territoriale e, history_toponyme t where
e.id_entite=t.id_entite and t.nom LIKE 'Cuisles' and code_insee_1999 LIKE
'51%'
```

#### Analyse de OUEZY :

L'entité avec ce nom ou ce code insee n'existe pas dans la BD Cassini\_EHESS.

```
select * from history_entite_territoriale e, history_toponyme t where
e.id_entite=t.id_entite and t.nom LIKE 'Ouezy' and code_insee_1999 LIKE
'14%'
```

#### Analyse de LIEOUX :

Pour **LIEOUX**, il existe pourtant bien une entité avec ce toponyme dans la base de données GeoPeuple, mais son code INSEE est différent (31483). Cette entité n'a pas de limites après l'importation. Elle est sensée avoir disparue d'après la BD Cassini\_EHESS lors d'une fusion en 1973.

```

select * from history_entite_territoriale e, history_toponyme t where
e.id_entite=t.id_entite and t.nom LIKE 'Lieoux' and code_insee_1999 LIKE
'31%'
select * from history_limite where id_entite=19433
select * from commune where code_insee_court LIKE 31483

```

### 5.2.2. Erreurs détectées dans les géométries anciennes

Certaines géométries ne sont pas correctes, voir Figure 13. Une diagonale à droite se dessine qui correspond aux géométries fausses (issues de Xfusion6.dat). Certaines des géométries de Xfusion6.dat sont correctes cependant.



**Figure 13. Géométries récupérées le 24 mai 2012 à EHESS.**

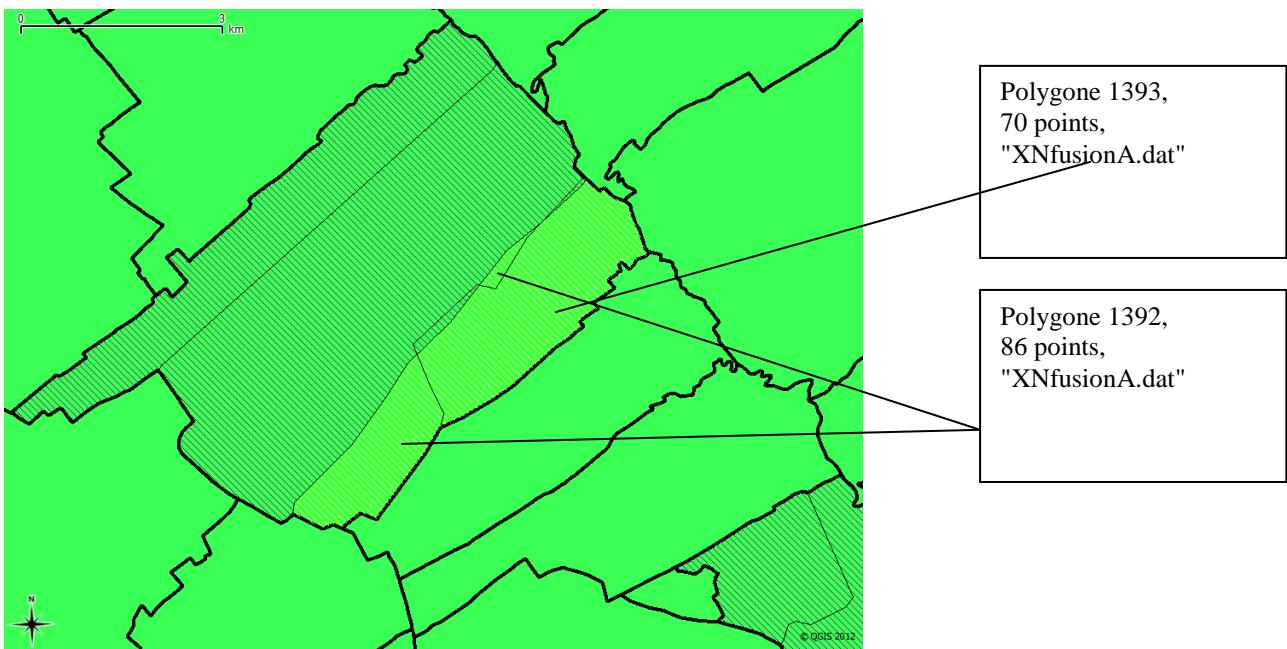
Il apparaît que la géométrie de Barberey-aux-Moines (numéro d'ordre = 2620) est incorrecte. En fait deux géométries ont été saisies, et celle qui est correcte correspond au polygone gid=1392, (le plus grand polygone n°1392 dans la figure 3).

```

select numordre, validdate, nom, code_insee_1999, code_insee_court,
c_suite_dev, c_suite_1790_dev
from limitesanciennescommunes, commune
where num_ordre=numordre and num_ordre=2620

Pour corriger :
update limitesanciennescommunes set validdate = '1791-1800' where gid=1392
delete from limitesanciennescommunes where gid=1393

```



**Figure 14. Barberey-aux-Moines : deux géométries, dont une incorrecte (1393). Les traits gras correspondent aux limites de la BD Carto actuelle, les traits fins aux hypothèses LeBras.**

Le Tableau 78 liste parmi les géométries livrées par Hervé LeBras celles qui sont incorrectes. Ce tableau rapporte le numéro de ligne (tel que spécifié dans le fichier Xfusion6.dat original) et le numéro d'ordre correspondant qui a été calculé. Le champ "validDate" renseigne sur la période de validité de la géométrie tel que donné dans Xfusion6.dat. Par conséquent, 1791 doit être compris comme 1793, et lorsqu'une seule date est donnée, elle correspond à la fin de validité de la géométrie.

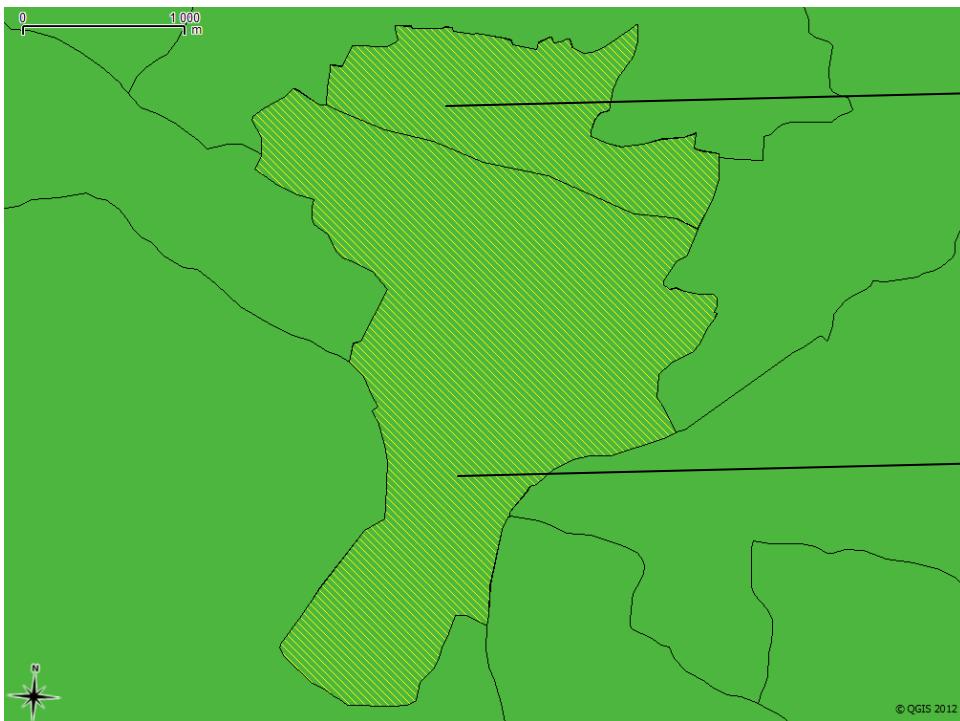
nbPoints	codeInsee	numLigne	numOrdre	validDate	nom
27	91016	42967	11993	1791-1974	Dommerville
81	90097	42956	41057	1791-1972	Vourvenans
99	90082	42940	12922	1791-1972	Eschêne-Autrage
77	90042	42898	35070	1791-1972	Salbert
51	90041	42896	13301	1791-1973	Étueffont-Bas
154	89472	42843	40702	1791-1972	La Villotte
145	89387	42750	29703	1791-1972	Rosoy
212	89345	42708	2235	1791-1970	Avrolles
88	89264	42628	40215	1791-1964	Villegardin
154	89246	42608	9556	1791-1964	Civry-sur-Serein
192	89200	42561	9350	1791-1947	Chichy
60	89169	42529	14057	1791-1971	La Chapelle-Vieille-Forêt
56	89086	42450	24099	1791-1943	La Mothe-aux-Aulnaises
111	89073	42437	20098	1791-1972	Louesme
51	89024	42385	39115	1791-1972	Vaux
80	89014	42374	11846	1791-1976	Dilo
91	89005	42363	11425	1791-1972	Cusy
132	88478	42295	16087	1791-1882	Graux
78	88393	42209	13199	1791-1905	L'Étanche
106	88372	42187	24870	1791-1946	Neuveville-lès-Raon
49	88305	42110	15775	1791-1965	Gouécourt

102	88274	42077	28883	1791-1962	Rémois
88	88227	42027	29565	1791-1977	Roncourt
76	88146	41939	15556	1791-1972	Girovillers-sous-Montfort
112	88114	41904	25956	1791-1964	Outrancourt
294	88106	41896	2580	1791-1995	Ban-sur-Meurthe
146	87180	41757	34302	1791-1972	Saint-Priest-le-Betoux
383	87092	41650	22505	1791-1973	Milhaguet
781	87085	41638	3278	1791-1962	Beaune-les-Mines
218	87027	41565	33788	1791-1973	Saint-Nicolas-Courbefy
689	87014	41551	24063	1791-1972	Morterolles-sur-Semme
152	86120	41324	34396	1791-1972	Saint-Rémy-en-Montmorillon
212	86092	41290	34477	1791-1971	Saint-Romain-sur-Vienne
121	86079	41276	5056	1791-1972	Le Bouchet
35	86066	41254	37094	1791-1972	Targé
67	86005	41185	30821	1791-1971	Saint-Cassien
101	85277	41143	34710	1791-1973	Saint-Sornin
73	85243	41114	33793	1791-1973	Saint-Nicolas-de-Brem
135	85158	41023	21458	1791-1964	Saint-Martin-sous-Mouzeuil
315	85094	40951	26340	1791-1968	Payré-sur-Vendée
213	85046	40896	19620	1791-1973	La Limouzinière
106	85040	40890	32314	1791-1974	Saint-Hilaire-du-Bois
35	85036	40884	9641	1791-1999	La Claye
498	83124	40643	6094	1791-1970	Brovès
155	81262	40158	6721	1791-1972	Campes
247	50629	23132	15405	1791-1972	Gerville-la-Forêt
106	50611	23113	24492	1791-1963	Nacqueville
76	50597	23096	31695	1791-1972	Sainte-Eugienne
150	50590	23087	25142	1791-1972	Noirpalu
159	50514	23015	34580	1791-1972	Saint-Sauveur-de-Chaulieu
80	50415	22905	36938	1791-1958	Sainte-Suzanne-en-Baupuis
80	50405	22885	18855	1791-1964	Lastelle
91	50218	22691	33812	1791-1962	Saint-Nicolas-près-Granville
129	50173	22638	16570	1791-1964	Hainneville
261	50144	22604	4630	1791-1850	Boisbenâtre
150	50109	22564	5254	1791-1964	Boureys
187	50031	22485	6997	1791-1964	Carteret
71	80703	39696	16329	1791-1972	Guémicourt
154	80623	39595	25971	1791-1972	Onvillers
45	80617	39583	28412	1791-1968	Le Quesnoy
44	80608	39572	17432	1791-1965	Hyencourt-le-Petit
220	80570	39513	7071	1791-1965	Castel
78	80566	39508	13848	1791-1974	Fieffes
<b>70</b>	<b>13349</b>	<b>10349</b>	<b>2620</b>	<b>1791-1800</b>	<b>Barberey-aux-Moines</b>

Tableau 78. Liste des géométries qui posent encore problème (non insérées en BD).

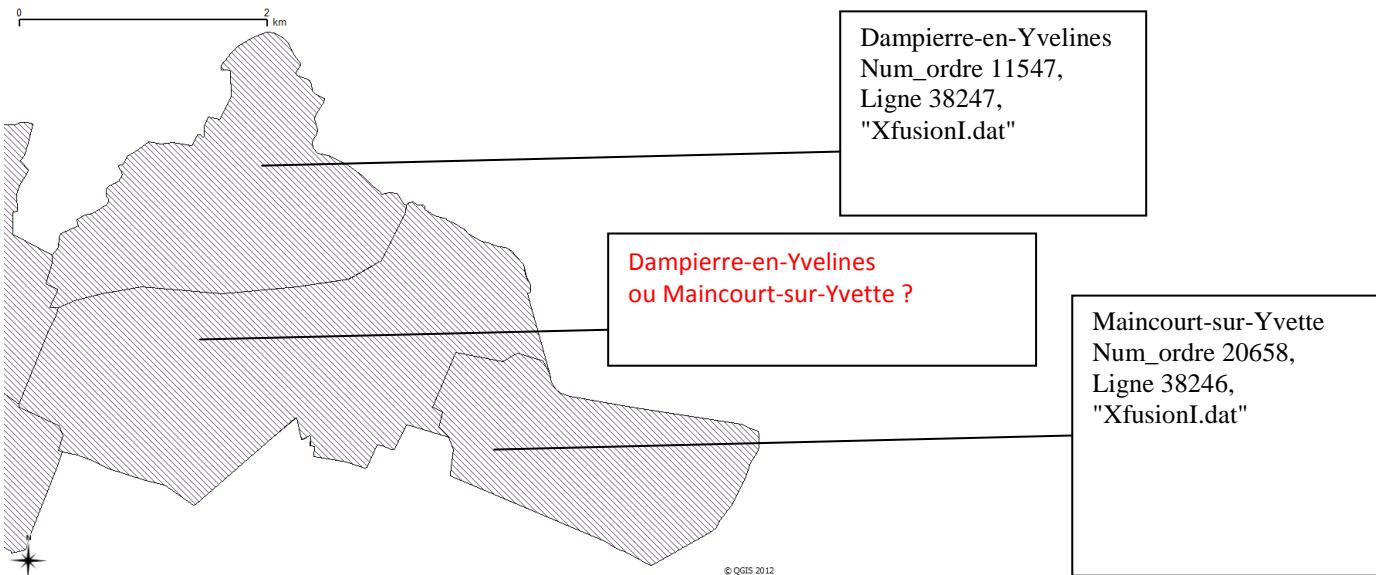
L'insertion des limites anciennes bute sur des doublons : parfois ce sont juste des géométries définies deux fois (après vérification manuelle, on est certain que ce ne sont que des doublons, comme dans

le cas « Valcongrain » et « Campandré » de la Figure 15). Toutes entrées des fichiers Xfusion4.dat sont identiques à celles de XfusionB.dat. De même pour XfusionF.dat et XfusionE.dat.



**Figure 15. Campandré-Valcongrain : doublons.**

Sinon, ce sont aussi des géométries mal définies comme dans le cas de Maincourt-sur-Yvette et Dampierre-en-Yvelines (Figure 16), dont le polygone central est défini 2 fois, une fois comme appartenant à Maincourt-sur-Yvette et une autre fois comme appartenant à Dampierre-en-Yvelines.



**Figure 16. Dampierre-en-Yvelines et Maincourt-sur-Yvette : appartenance du polygone central ?**

## 6. Bibliographie

- Chareille P., Rodier X. & Zadora-Rio E. Analyse des transformations du maillage paroissial et communal en Touraine à l'aide d'un SIG. *Histoire & mesure*, vol. 19, no. 3/4, pages 317–344, 2004.
- Clifford J., Dyreson C.E., Isakowitz T., Jensen C.S., and Snodgrass R.T.. "On the Semantics of Now in Databases." *ACM Transactions on Database Systems* 22(2), 1997, pp. 215—254
- Commission sociale des évêques de France, « Église et société face à l'aménagement du territoire », Bayard Éditions, 1998, pp. 157-158.
- Guibert B., Laganier J., Volle M. Essai sur les nomenclatures industrielles. *Economie et Statistique*, n°20, 1971
- Lang G., Le code officiel géographique (COG). Avant, pendant et autour. Direction de la coordination statistique et des relations internationales, INSEE, 2010, accessible sur [http://projetbabel.org/gl/cog\\_index.htm](http://projetbabel.org/gl/cog_index.htm)
- Motte C., Séguin I., Théré C. Communes d'hier, communes d'aujourd'hui. Les communes de la France métropolitaine, 1801-2001. Dictionnaire d'histoire administrative. Institut National d'Études Démographiques, 2003.
- Motte C., Péliquier J.-P. Géonoméclature historique des lieux habités, Direction des archives de France, 2003.
- Motte C., Vouloir M-C., « Frontières administratives et identités communales. Le cas de la France, XVIIIe-XXe siècles », *The Historical Review*, Athènes, vol. V, 2008