

pour comprendre le présent et construire un avenir durable

Juin 2012

Version 1.1



Fichiers fonciers 2009 et 2011 Principes de traitement, outils choisis et descriptif de livraison

Sommaire

Principes généraux du traitement.....	3
Pré-requis à la lecture du document.....	3
Principales difficultés du traitement.....	3
Le choix d'un système de gestion de bases de données.....	3
La solution technique retenue pour le traitement.....	4
Le choix du SGBD PostgreSQL	4
Les 4 grandes étapes du traitement.....	4
Les données annexes produites.....	7
Descriptif de livraison des fichiers fonciers.....	8
Contenu du disque dur.....	8
Éléments de volumétrie.....	8

Version	Date	Auteurs	Commentaire
1	27 avril 2012	Rémi BOREL	Document de base
1.1	1 juin 2012	Rémi BOREL	Relecture Jérôme DOUCHÉ, mise en forme et prise en compte des versions

Principes généraux du traitement

Pré-requis à la lecture du document

Afin de bien comprendre la méthodologie de traitement, il faut connaître la **structure** et le **contenu** des fichiers sources.

Un document est disponible dans la rubrique « documentation générale sur les fichiers fonciers » du portail intranet « fichiers fonciers ».

Principales difficultés du traitement

Historiquement, les fichiers fonciers sont extrêmement **difficiles à traiter** pour des personnes non habituées à manipuler les outils informatiques. Cela est dû à 2 facteurs principaux.

Un volume très important

La principale difficulté du traitement national des fichiers fonciers est la gestion de leur **volume**. Mis bouts à bouts, les 600 fichiers d'un seul millésime représentent :

- 1 milliard de lignes,
- 130 Go de données.

Un format non standard

Le format des fichiers fonciers n'est compatible avec **aucun logiciel standard** de type tableur, base de données ou SIG.

De fait, il faut les **modifier** avant de pouvoir les intégrer dans un de ces logiciels.

Le choix d'un système de gestion de bases de données

Les outils bureautiques habituels comme OpenOffice, Office, Access ou Mapinfo ne sont pas adaptés.

Pourquoi pas des fichiers Mapinfo ?

Le format Mapinfo permet de gérer des données volumineuse. Néanmoins, ce format n'est **pas adapté** car :

- tous les chargés d'études ne disposent de Mapinfo,
- certaines tables des fichiers fonciers ne sont pas géographiques,
- la table des parcelles contient deux géométries (polygone et point).

Pourquoi un format de base de données ?

Le format « base de données » permet de disposer à la fois de données **géographiques** et non géographiques. Il permet également de gérer un **volume quasi illimité**.

Comment accompagner les utilisateurs novices ?

Les utilisateurs « **novices** » peuvent choisir :

- d'exporter rapidement les données dans des formats habituels,
- de tester les nombreuses fonctionnalités du SGBD¹.

L'utilisation d'un format de base de données permet aux **utilisateurs expérimentés** d'effectuer des traitements complexes et rationaliser l'espace disque.

La documentation écrite par le CETE Nord Picardie a pour objectif d'accompagner ces deux profils d'utilisateurs.

1 SGBD : Système de Gestion de Bases de Données

La solution technique retenue pour le traitement

Le choix du SGBD PostgreSQL

PostgreSQL est un **système de gestion de base de données** comparable à Access ou Oracle.

Particularités et atouts

Les principaux atouts de PostgreSQL sont :

- sa licence libre,
- sa robustesse,
- sa capacité à gérer des volumes de données quasi infinis.

La gestion des données géographiques

Il est possible d'installer des modules additionnels dans PostgreSQL. Il existe en particulier un **module spécifique « PostGIS »** permettant de gérer les données géographiques.

Le couple PostgreSQL / PostGIS devient alors un véritable outil de traitement de données géographiques permettant :

- d'importer des données,
- d'effectuer des traitements géométriques comme des :
 - zones tampons,
 - calculs d'intersection,
 - requêtes spatiales, etc,
- d'afficher les données dans des logiciels SIG comme QGIS,
- d'exporter les données.

Les 4 grandes étapes du traitement

Afin d'éviter aux services ce travail fastidieux, les fichiers fonciers traités par le CETE Nord Picardie ont un **format standardisé**. Les traitements sont également **identiques** sur la France entière. Cela permet de réaliser des **comparaisons homogènes** d'un territoire à l'autre.

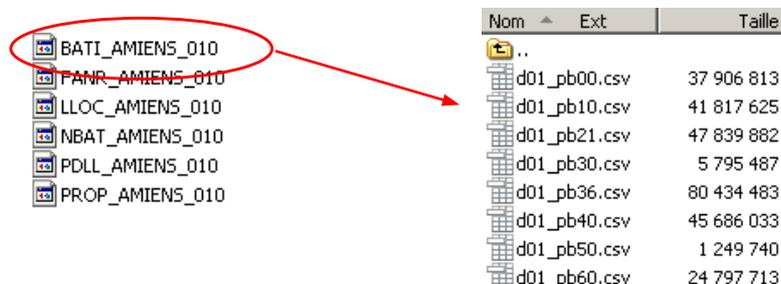
Pour cela, le traitement suit 4 étapes.

Étape 1 : transformation des fichiers bruts en CSV

La première étape consiste à créer des fichiers CSV à partir des 6 fichiers sources. Chacun de ces 6 fichiers engendrent **autant de fichiers CSV que d'articles qu'ils contiennent**.

Le découpage est effectué par des **scripts** développés en Python, combinés à des **requêtes SQL**. Le traitement est réalisé direction fiscale par direction fiscale.

Par exemple, le fichier des propriétés bâties engendre 8 fichiers CSV.



Nom	Ext	Taille
..		
d01_pb00.csv		37 906 813
d01_pb10.csv		41 817 625
d01_pb21.csv		47 839 882
d01_pb30.csv		5 795 487
d01_pb36.csv		80 434 483
d01_pb40.csv		45 686 033
d01_pb50.csv		1 249 740
d01_pb60.csv		24 797 713

Découpage du fichier des propriétés bâties 2011 du département de l'Ain

Les fichiers CSV ainsi obtenus peuvent ensuite être intégrés dans PostgreSQL.

La solution technique retenue pour le traitement (suite)

Les 4 grandes étapes du traitement (suite)

Étape 2 : intégration des fichiers CSV dans PostgreSQL

Cette étape est réalisée grâce à des **requêtes SQL** standard. On obtient ainsi des tables départementales issues des 6 fichiers.

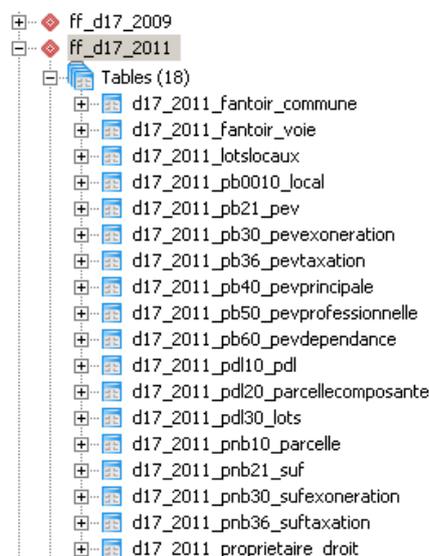
Millésime	Nombre de tables par département
2009	19 (une table en plus due au changement de version de MAJIC)
2011	18

Étape 3 : réalisation des traitements

Les traitements sont également réalisés grâce à des requêtes SQL. Les grands principes de ce traitement sont :

- la création des tables dans lesquelles les champs indisponibles ont été supprimés, et les **identifiants** ajoutés,
- la création de tables temporaires contenant les nouveaux **indicateurs**. Cette étape est possible grâce à la présence des identifiants créés précédemment,
- la combinaison des tables initiales et des tables temporaires qui donne les tables **finales**.

Dans PostgreSQL, les données sont organisées par **schéma**. Un schéma est



Le schéma « **ff_d17_2011** » contient les fichiers fonciers 2011 pour le département 17.

Pour les 100 départements français, cela représente :

- 190 tables et 100 schémas pour 2009,
- 180 tables et 100 schémas pour 2011.

Le nom du schéma contient toujours les références au millésime et au département : **ff_d<XXX>_<AAAA>**
Avec <XXX> : code département
<AAAA> : millésime

Organisation des tables finales dans PostgreSQL

La solution technique retenue pour le traitement (suite)

Les 4 grandes étapes du traitement (suite)

Les tables des fichiers fonciers respectent également une **convention** de nommage : d<XXX>_<AAAA>_<fichier><article>_<nom-article> avec :

- <XXX> : code département,
- <AAAA> : millésime,
- <fichier> : nom du fichier d'origine dans les fichiers fonciers bruts,
- <article> : numéro d'article du fichier brut,
- <nom-article> : nom de l'article.

Exemple : la table « d972_2009_pnb21_suf » est la table des subdivisions fiscales (**suf**) de la Martinique (**972**) pour le millésime **2009**. Cette table vient de l'article **21** du fichier des propriétés non bâties (**pnb**).

Les tables « lotslocaux » et « propriétaire_droit » n'ont pas de numéro d'article car le fichier source n'en contient qu'un seul.

La table « pb0010_local » est une réunion des articles 00 et 10 du fichier des propriétés bâties (pb).

Étape 4 : export des données

Les données des fichiers fonciers sont **exportées schéma par schéma** à l'aide de l'utilitaire « pg_dump ».

Ainsi, il existe un fichier de données (fichier .dump) par département et par année. Ces fichiers contiennent une série de requêtes SQL compressées qui permettent de **reconstituer** les données dans PostgreSQL.

Nom	Taille
ff_d02_2009.dump	321 449 Ko
ff_d02_2011.dump	307 856 Ko
ff_d60_2009.dump	329 135 Ko
ff_d60_2011.dump	317 570 Ko
ff_d80_2009.dump	234 652 Ko
ff_d80_2011.dump	224 887 Ko

Exemple de fichiers exportés

Chaque fichier dump est **nommé** exactement comme le schéma des données qu'il contient.

La solution technique retenue pour le traitement (suite)

Les données annexes produites

À partir des fichiers fonciers traités, le CETE Nord Picardie a construit des **tables annexes** généralisées à diverses échelles.

Ces tables contiennent des indicateurs pour la plupart issus de la table des **parcelles**.

Données agrégées à des niveaux administratifs

Les niveaux administratifs couverts, de la plus grande échelle à la plus petite :

- section cadastrale,
- commune,
- canton,
- arrondissement,
- département,
- région.

Table des bâtiments

Dans les fichiers fonciers, la notion de bâtiment existe en tant que **lettre de bâtiment** d'un local. La table des bâtiment est donc l'agrégation des locaux ayant la même lettre de bâtiment.

Données carroyées

Les données « carroyées » sont indépendantes de tout découpage administratif. Il s'agit de représenter les informations sur des **carrés** de 100m, 1km ou 10km de côté.

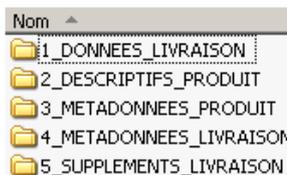
Descriptif de livraison des fichiers fonciers

Contenu du disque dur

Les fichiers fonciers 2009 et 2011 ont été livrés en mai 2012 par le PNE **référentiels**. Le disque dur envoyé aux services contient également les référentiels IGN habituels fournis régulièrement par le PNE.

Une arborescence habituelle

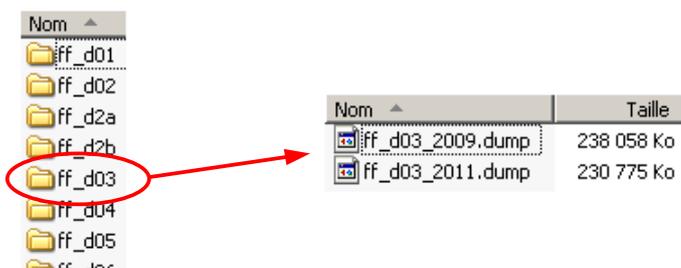
L'arborescence des données livrées est similaire à celle des référentiels IGN.



Dossiers contenant les données

Dossier 1 : données

Le dossier 1 contient les **fichiers fonciers traités** par le CETE Nord Picardie. Les données sont classées **par département**. Chaque sous-dossier contient les fichiers fonciers 2009 et 2011.



Contenu des dossiers des fichiers fonciers

Dossiers 2, 3 et 4 : la documentation

Les dossiers 2, 3 et 4 contiennent seulement :

- la déclaration CNIL,
- un lien vers le portail fichiers fonciers.

En effet, la documentation a pour vocation à être régulièrement **mise à jour** sur intranet.

Dossier 5 : suppléments

Le dossier 5 contient les données **agrégées** :

- tables des bâtiments,
- tables des carroyages,
- tables des données agrégées à différents niveaux administratifs.

Pour tous les services, ces données sont livrées sur **l'ensemble du territoire** français.

Éléments de volumétrie

Les fichiers dump contenus dans le disque dur sont des « images » **compressées** de données.

Une fois ces fichiers importés dans PostgreSQL, le volume réel des données doit être **multiplié par 10**.

Par exemple, le fichier « ff_d29_2011.dump » fait 630 Mo. Les données finales occuperont un espace de **6 à 7 Go**.