RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Cerema
CLIMAT & TERRITOIRES DE DEMAIN

Documentation des Données foncières

Extrait de la documentation en ligne



DOCUMENTATION

MINISTÈRE
DE LA TRANSITION
ÉCOLOGIQUE
ET DE LA COHÉSION
DES TERRITOIRES
Liberté
Égalité
Fraternité

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Cerema
CLIMAT & TERRITOIRES DE DEMAIN

Le Cerema est un établissement public sous la tutelle du ministère de la Transition écologique, présent partout en métropole et dans les Outre-mer grâce à ses 26 implantations et ses 2 400 agents. Détenteur d'une expertise nationale mutualisée, le Cerema accompagne l'État et les collectivités territoriales pour la transition écologique, l'adaptation au changement climatique et la cohésion des territoires par l'élaboration coopérative, le déploiement et l'évaluation de politiques publiques d'aménagement et de transport. Doté d'un fort potentiel d'innovation et de recherche incarné notamment par son institut Carnot Clim'adapt, le Cerema agit dans 6 domaines d'activités : Expertise & ingénierie territoriale, Bâtiment, Mobilités, Infrastructures de transport, Environnement & Risques, Mer & Littoral.

Site web : www.cerema.fr

Avertissement

Ce document est extrait de la documentation en ligne, disponible sur le site https://doc-datafoncier.cerema.fr/, qui est mise à jour en continu. Ce document n'est donc valable qu'à un instant t, et le lecteur est invité à aller consulter la documentation en ligne pour bénéficier des dernières nouveautés.

Ce guide est un guide technique renseignant sur le contenu, les limites et la manière d'utiliser la base de données « Fichiers fonciers ». Pour d'autres questions (modalités d'accès, études réalisées…), le lecteur est invité à consulter le site https://datafoncier.cerema.fr

Cette documentation, évolutive, peut cependant encore contenir des erreurs ou coquilles. Si vous relevez une erreur dans celle-ci, n'hésitez pas à nous en faire part par mail via la boite datafoncier@cerema.fr

Table des matières

- De DVF à DV3F
  - Objectifs
  - Rédacteurs
  - Licence
  - DVF pour Demande de Valeurs Foncières
  - Le modèle DVF+
  - Pour compléter votre lecture...
  - Le modèle DV3F

De DVF à DV3F

Objectifs

Ce guide a pour objectifs :

- de présenter ce que sont DVF, DVF+ et DV3F,
- de proposer une prise en main au regard des principaux usages,
- de proposer une description des notions essentielles autour de DV3F,
- d'expliciter les méthodes de construction de la donnée.

Elles contiennent de véritables bases pour débuter avec DVF+/DV3F.

Rédacteurs

Ce guide a été rédigé par Magali Journet, Antoine Herman et Olivier Dussart (Cerema).

Dernière mise à jour : juin 2021

Licence

Tous les contenus proposés sont sous licence Creative Commons CC-BY-NC-SA.

DVF pour Demande de Valeurs Foncières

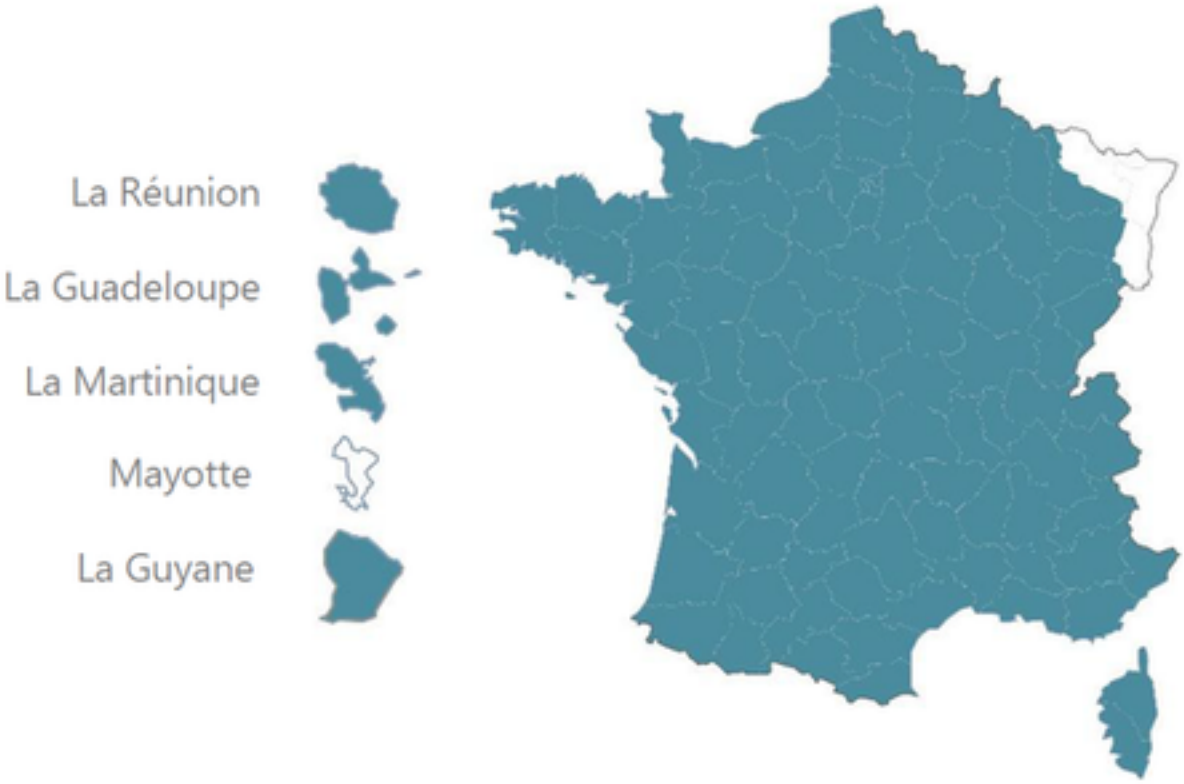
DVF est une base de données fiscale produite par la Direction Générale des Finances Publiques (DGFiP). Elle permet l'obtention gratuite des données présentant les valeurs foncères de l'ensemble des **transactions immobilières et foncières à titre onéreux des 5 dernières années** ainsi que de certaines de leurs caractéristiques :

- nature de la vente (vente, adjudication, VEFA, vente de TAB…);
- localisation des biens (références cadastrales, adresses…);
- surface des parcelles et des locaux concernés;
- types de locaux (maisons, appartements, dépendances, locaux industriels ou commerciaux).

Depuis avril 2019 et suite au décret du 28 décembre 2018 relatif à la publication des informations portant sur les valeurs foncières, **les données DVF ont été ouvertes à l'ensemble du public**. Les données sources sont disponibles sur le site data.gouv.fr. Les différentes évolutions législatives liées à la mise à disposition gratuite des données sur les transactions immobilières s'inscrivent dans une démarche croissante de l'administration française au service d'une meilleure transparence des marchés et d'une meilleure gestion des territoires.

Périmètre géographique de DVF

En terme de périmètre géographique, DVF est disponible pour la **France métropolitaine**, hors Alsace et Moselle, **ainsi que les DROM**, excepté Mayotte.



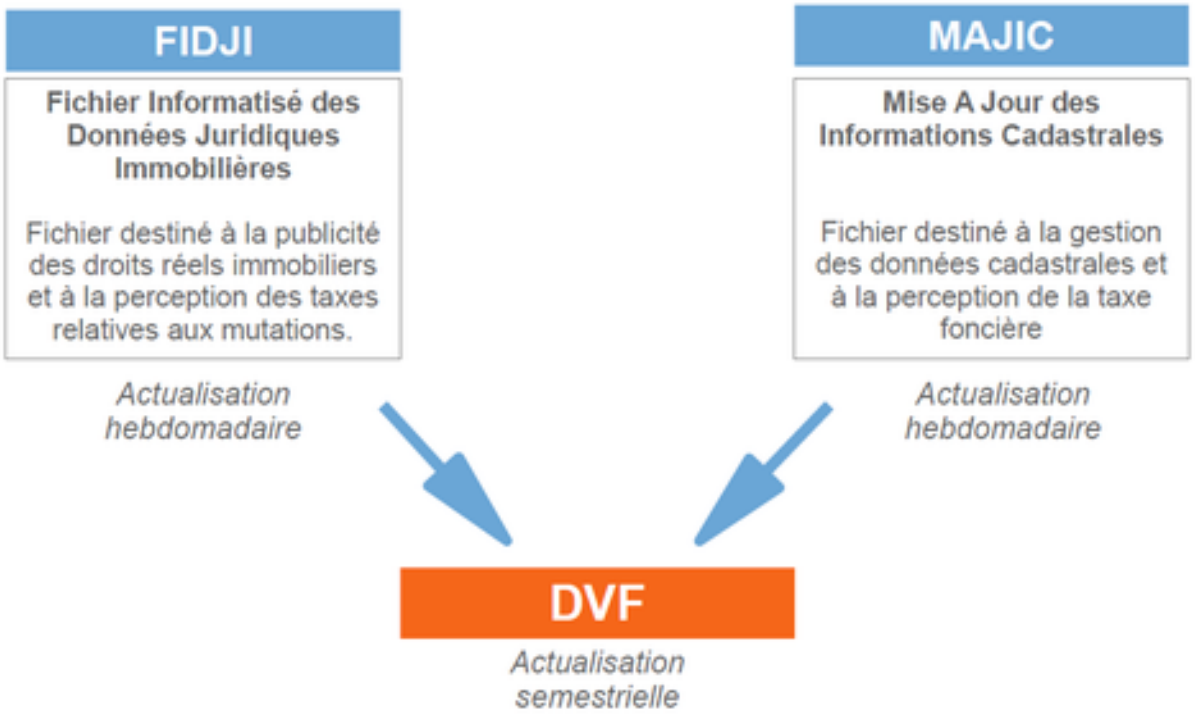
(source GnDVF)

Constitution de DVF

Des données issues de FIDJI et MAJIC

DVF est constituée à partir de 2 sources de données fiscales produites par la DGFiP.

Le traitement DVF récupère, à la fois, des informations renseignées par les services de la publicité foncière dans la base de données FIDJI (Fichier Informatisé des Données Juridiques Immobilières) et des informations issues de la base de données MAJIC (Mise à Jour des Informations Cadastrales) pour produire des fichiers actualisés et accessibles à un rythme semestriel.



Précisions sur les mutations présentes dans DVF

DVF regroupe toutes les mutations onéreuses en France, hors Alsace, Moselle, Mayotte à trois exceptions :

- les mutations non onéreuses : succession, legs, etc.;
- les mutations ne passant pas par un notaire : les ventes de parts de SCI, certaines ventes de public à public, etc;
- les mutations dites « complexes » par la DGFiP. Celles-ci correspondent en général soit à des parcelles ayant de très grandes copropriétés ou bien avec un litige non réglé.

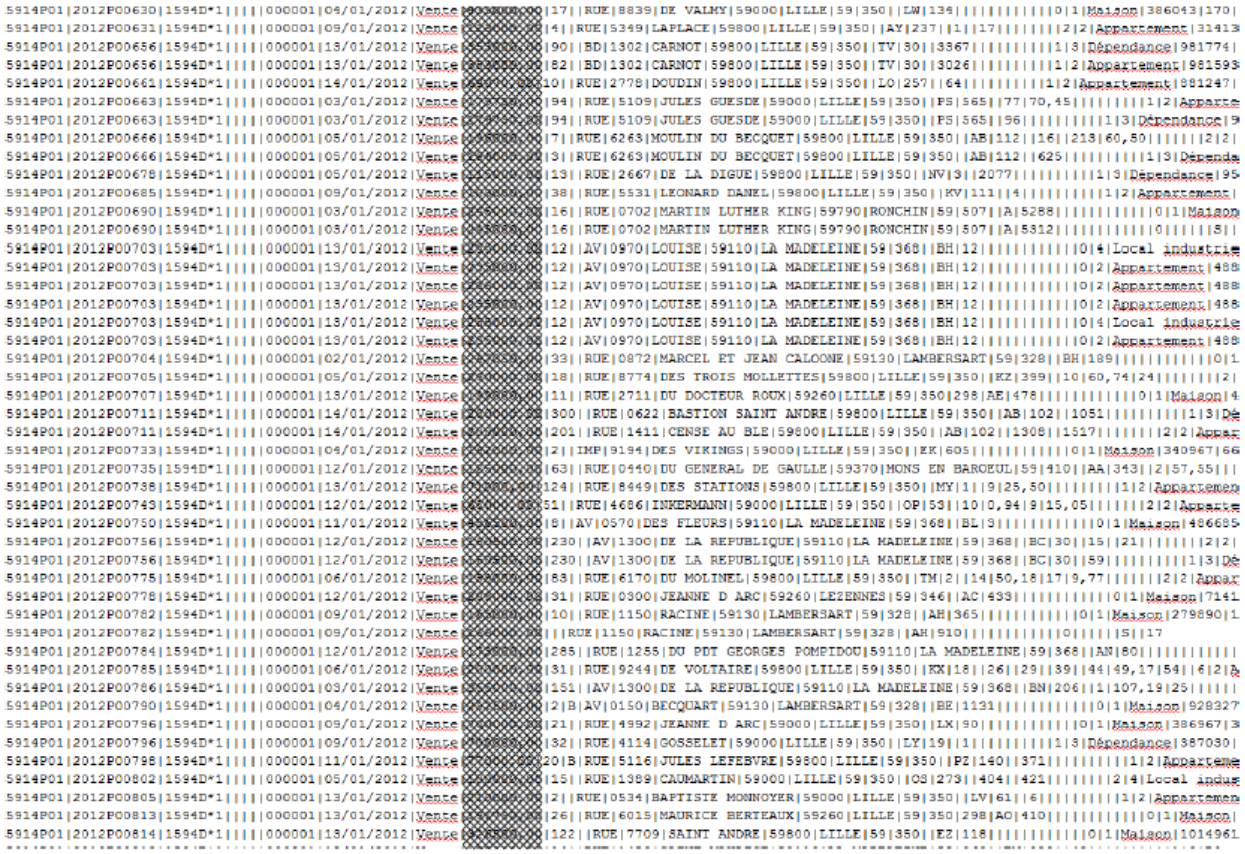
Une donnée brute en 43 champs

La donnée DVF livrée se présente sous la forme de fichiers textes, encodés en UTF-8, lisibles sous différentes éditeurs de textes ou tableurs.

Les différentes lignes de ces fichiers sont composées de 43 champs séparées par le caractère "|". Elles décrivent tout ou partie des éléments d'information d'une transaction immobilière (ou mutation).

A noter que les données DVF ouvertes et les données DVF "historiques" (avant la mise en open-data) sont structurées de la même manière mais certains champs ont été masqués (cf tableau ci-après).

L'illustration ci-dessous présente un échantillon d'un fichier source DVF :



Les 43 champs présents dans DVF sont les suivants :

Libellé	Exemple	Observation
Code du service CH	5914P01	masquée depuis l'ouverture publique de la donnée
Référence document	2009P00999	masquée depuis l'ouverture publique de la donnée
1er Article CGI	1594D*1	masquée depuis l'ouverture publique de la donnée
2e Article CGI	257-7-1*2	masquée depuis l'ouverture publique de la donnée

3e Article CGI		masquée depuis l'ouverture publique de la donnée
4e Article CGI		masquée depuis l'ouverture publique de la donnée
5e Article CGI		masquée depuis l'ouverture publique de la donnée
Numéro de disposition		
Date de mutation	16/01/2010	
Nature de mutation	Vente	
Valeur foncière	200 000 €	
Numéro de voie	01	
B/T/Q	T	
Type de voie	Rue	
Code voie	4242	
Voie	Jean Jaures	
Code postal	59000	
Commune	LILLE	
Code du département	59	
Code de la commune	350	
Préfixe de section	298	
Section	AB	
Numéro de plan	100	
Numéro de volume	-	
1er lot	-	

Surface Carrez 1er lot	-	
2e lot	-	
Surface Carrez 2e lot	-	
3e lot	-	
Surface Carrez 3e lot	-	
4e lot	-	
Surface Carrez 4e lot	-	
5e lot	-	
Surface Carrez 5e lot	-	
Nombre de lots	-	
Code du type de local	2	
Type du local	Maison	
Identifiant du local	999999	masquée depuis l'ouverture publique de la donnée
Surface réelle bâti	88	
Nombre pièces princ	4	
Nature de culture	S	
Nature de culture spéc.		
Surface de terrain	140	

Atouts et Limites de DVF pour l'observation foncière et immobilière

Atouts principaux

Les données DVF présentent plusieurs atouts. Elles sont :

- disponibles à une échelle nationale,
- gratuites et désormais libres,
- réputées exhaustives,
- globalement fiables,
- facilement géolocalisables.

Limites principales

Les données DVF présentent toutefois quelques limites :

- Elles sont complexes à exploiter en l'état,
- Elles peuvent être difficilement intelligibles pour les novices,
- Elles n'ont pas vocation première à permettre l'observation foncière et les informations fournies apparaissent rapidement limitantes pour une analyse fine.

La mise en place de modèles de données tels que DVF+ et DV3F cherchent à repousser ces limites en proposant notamment une structuration et un enrichissement de la donnée source.

Le modèle DVF+

Contexte

Les travaux autour de la création d'un modèle permettant de faciliter le passage des données DGFiP, « Demande de valeurs foncières » (DVF), en base de données aisément exploitable s'inscrivent dans un processus qui a débuté en 2011 à l'initiative d'un groupe technique lancé par l'ADEF, aujourd'hui devenu le Groupe National DVF (GnDVF).

Ce groupe a associé, dès 2013, le CETE Nord-Picardie (aujourd'hui Direction Hauts-de-France du Cerema) afin de produire un modèle de données pour l'import des données natives de DVF partagé.

Le Cerema a ensuite travaillé, pour le compte de la DGALN, à étoffer ce modèle en y ajoutant de nouvelles variables au sein d'un premier modèle, baptisé DVF+.

Aujourd'hui, et depuis l'ouverture des données DVF en 2019, le Cerema et la DGALN mettent à disposition les données DVF préparées sous ce modèle DVF+. Ces données sont téléchargeables sur le site Datafoncier.

Intérêt d'une structuration en base de données de DVF

Les données DVF brutes fournies par la DGFiP sont livrées sous un format texte lisible depuis un éditeur de texte ou un tableur classique. Chaque ligne du fichier source est souvent une décomposition partielle d'une vente et comporte 43 champs qui mélangent des informations liées à différentes notions fiscales. Il devient très vite difficile de travailler à une analyse sérieuse à partir d'un tel fichier.

La mise en place d'une structuration de la donnée est vite nécessaire pour travailler efficacement sur cette donnée et l'enrichir plus aisément.

En terme de structuration de la donnée initiale, chaque table du modèle partagé DVF+ représente une unique notion de DVF à laquelle sont affectées les données qui y sont directement liées. Cette structuration permet de supprimer les données dupliquées et de mettre en place des liaisons plutôt intuitives entre les tables.

Par ailleurs, le modèle DVF+ ajoute, à une donnée d'origine fiscale, des variables issues de ces même données mais transformées ou agrégées pour permettre une exploitation plus aisée pour l'observation foncière et immobilière.

Dans la version DVF+ - open-data proposée depuis l'ouverture des données d'avril 2019, différents champs géométriques issus du PCI Vecteur sont également intégrés pour offrir une base de données géolocalisée.

Tables constitutives du modèle

Le modèle DVF+ est composé de 12 tables (dont 3 principales) et de 5 tables annexes.

Les 12 tables principales ou secondaires sont les suivantes :

Nom de la table	Description
mutation	**Table des mutations**
mutation_art_cgi	Table des articles du code général des impôts attachés à la mutation
disposition	Table des dispositions
disposition_parcelle	**Table des parcelles attachées à la disposition**

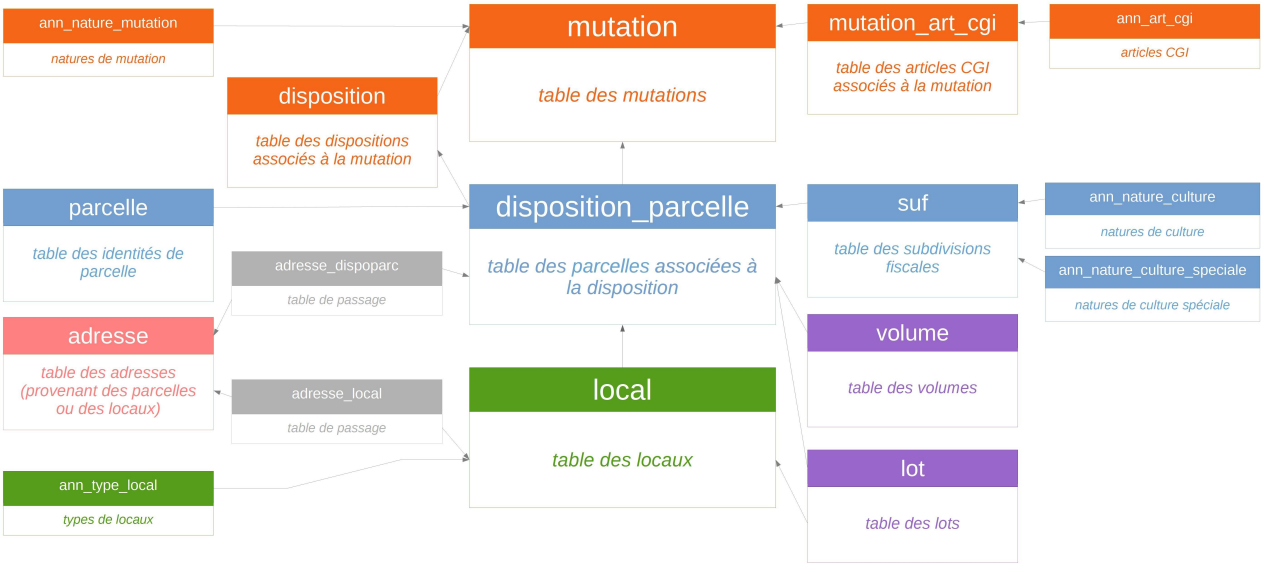
parcelle	Table des parcelles
local	**Table des locaux**
suf	Table des subdivisions fiscales
volume	Table des volumes
lot	Table des lots
adresse	Table des adresses
adresse_dispoparc	Table de liaison entre la table adresse et la table disposition_parcelle
adresse_local	Table de liaison entre la table adresse et la table local

Les tables mutation, disposition_parcelle et local sont dites "principales" dans le sens où les principales variables ont y été recentrées afin de faciliter l'utilisation du modèle par des utilisateurs peu aguerris dans l'exploitation de bases de données. La grande majorité des analyses liées à l'observation foncière et immobilière peut être réalisée à partir des données issues de ces trois tables.

Les 5 tables annexes sont les suivantes :

Nom de la table	Description
ann_nature_mutation	Table contenant les natures de mutation
ann_cgi	Table contenant les articles CGI
ann_type_local	Table contenant les types de locaux
ann_nature_culture	Table contenant les natures de culture
ann_nature_culture_speciale	Table contenant les natures de culture spéciales

Un schéma simplifié du modèle est présenté ci-dessous :



Variables constitutives du modèle

Des variables brutes et des variables plus élaborées facilitant l'analyse

Plusieurs catégories de variables existent au sein de "DVF+":

Catégorie	Description	Exemples
Identifiants de clef primaire	Entiers ou chaînes de caractère permettant de désigner une et une seule entité d'une table. L'identifiant de clef primaire de la table apparait au début de chacune des tables.	idmutation, iddispo, iddispoloc, etc.
Données brutes	Données provenant des données sources de la DGFiP et intégrées sans transformation dans le modèle.	refdoc, coddep, scarrez, etc.
Données brutes transformées	Données provenant des données sources de la DGFiP et intégrées sous une forme transformée (par concaténation, découpage, etc.).	anneemut, idpar, idloc, etc.
Données calculées et/ou agrégées	Données issues de calculs (mathématique, conditionnel, etc.) ou d'agrégations effectués à partir d'autres données du modèle.	vefa, nbsufidt, dcnt01, nbparmut, etc.

Chacune des variables du modèle DVF+ est calculée uniquement à partir des données brutes de DVF, ces dernières étant d'ailleurs intégralement restituées par le modèle. Il n'y a pas de données exogènes à ce stade. Les variables calculées visent à pouvoir s'appliquer sur l'ensemble du territoire et à relever d'une méthodologie ne pouvant être remise en cause.

Ces variables restent plutôt intitutives pour correspondre à des éléments d'analyse représentatifs des besoins courants des utilisateurs. Elle permettent, à la fois, d'effectuer rapidement des calculs d'indicateurs mais aussi d'aboutir à la mise en œuvre de méthodologies locales ou nationales plus complexes (ex: typologie de bien).

Syntaxe des variables

Quelques règles syntaxiques ont été mises en place sur les variables pour donner quelques repères à l'utilisateur :

Si la variable commence par...	Alors il s'agit d'un(e)...
id	identifiant
cod	code (postal, insee, etc...)
s	surface
nb	denombrement
l_	liste (type array)
lib	libellé

A noter que toutes les variables comportent au maximum 10 caractères de manière à faciliter l'exportation au format shp.

Pour compléter votre lecture...

A savoir qu'un guide d'accompagnement sur DVF, co-construit par une large communauté d'utilisateurs, existe. Il est disponible ou téléchargeable sur le site du Groupe national DVF.

Une littérature détaillée concernant les Fichiers fonciers est également disponible sur le site Datafoncier

Le modèle DV3F

Contexte

Une fois le travail de structuration de la donnée accomplie avec le modèle DVF+, et toujours dans l'objectif de faciliter l'observation foncière et immobilière, la DGALN a demandé au Cerema d'engager un travail d'enrichissement de la donnée DVF à partir de données exogènes.

Ainsi, une étude expérimentale réalisée en 2015 et financée par l'EPF Nord Pas-de-Calais et la DGALN, a permis de mesurer la faisabilité d'un croisement de DVF avec les Fichiers fonciers. Les résultats de cette étude, très encourageants, ont posé les bases de la construction d'un modèle de données dit "DV3F", reproductible à l'échelle nationale, qui élargit fortement les possibilités d'analyse. Ce modèle de données est en développement depuis mi-2015.

Ce modèle de données DV3F permet désormais de produire et de diffuser une base de données nationale gratuite, actualisée au moins une fois par an, sans réel équivalent pour l'analyse des marchés fonciers et immobiliers. Il fait l'objet d'améliorations continues issues des différents retours d'expérience des utilisateurs et validées par un Comité Technique (COTECH).

Conditions d'obtention de DV3F

Les conditions d'accès à DV3F sont, depuis 2020, communes à celles des Fichiers fonciers. La procédure d'obtention est décrite sur le site Datafoncier.

Les principaux apports de DV3F

Le modèle DV3F a d'abord pour objectif d'enrichir la donnée source DVF avec les variables des Fichiers fonciers intéressantes pour l'observation. Cet enrichissement est possible par le lien évident qui existe entre ses deux sources de données d'origine fiscale ; en effet, la donnée DVF provient des bases de données FIDJI et MAJIC de la DGFiP tandis que les Fichiers fonciers proviennent quasi-exclusivement de MAJIC. Les identifiants des différents biens sont communs et facilitent les rapprochement.

Le principal frein à ses rapprochements réside dans le fait que les Fichiers fonciers font état d'un "stock" au 1er janvier de l'année tandis que les Demandes de Valeurs Foncières retranscrivent plutôt le "flux" des ventes.

Une caractérisation plus fine des biens

Les caractéristiques des biens renseignés dans la donnée source DVF sont peu détaillées. Pour les locaux notamment, DVF fournit uniquement le type du local, la surface du bâti et le nombre de pièces principales.

Avec DV3F et les variables rapatriées des Fichiers fonciers, les informations sur les biens échangés sont plus précises et permettent ainsi d'aller plus loin dans la caractérisation de certains biens. A titre d'exemple, pour les locaux, DV3F permet d'obtenir l'année de construction, de préciser le nombre et les types de dépendance ou encore de mieux identifier l'usage des locaux.

La géolocalisation des données

Un des principaux intérêts de DV3F réside également dans la géolocalisation des données sources. L'apport de la géométrie parcellaire accroit fortement les potentiels d'usage : représentations cartographiques, croisements avec d'autres périmètres (PLU par exemple), requêtes géométriques variées, etc.

Les géométries proviennent d'un historique de parcelles constitué majoritairement à partir de différents millésimes de la BD Parcellaire (IGN) ainsi que du PCI Vecteur (ouvert au public). Cela permet, dans DV3F, de géolocaliser des parcelles qui auraient disparu à la suite d'une vente et ainsi d'obtenir des taux de géolocalisation très satisfaisants.

Une information sur les acheteurs et les vendeurs

DV3F vise à pallier l'absence d'informations sur les acquéreurs et vendeurs dans DVF. En travaillant à partir de plusieurs millésimes des Fichiers fonciers, il est en effet possible de retrouver le propriétaire d'un bien avant et après une vente et ainsi d'en déduire le vendeur et l'acheteur.

Les informations sur les propriétaires pouvant être récupérées à partir des millésimes disponibles des Fichiers fonciers sont donc rapatriées dans DV3F et accompagnées d'une estimation de la fiabilité liée au rapatriement.

Ces informations sont désormais anonymisées pour les personnes physiques et morales (suite à l'évolution des CGU depuis l'ouverture de DVF). Des catégories d'acteurs sont néanmoins proposées pour permettre des analyses sur certains segments de marchés.

Une contextualisation des transactions

Les différents millésimes des Fichiers fonciers fournissent une information annuelle de l'évolution des biens sur le territoire. Ainsi, par les croisements avec les Fichiers fonciers, DV3F permet, lorsque le recul temporel est suffisant, d'identifier la destination des biens vendus. Ces éléments de contexte de la vente, observés a posteriori, apportent des

informations précieuses pour certaines analyses (constructions ou démolition de locaux suite à la vente, aménagement de parcelles, stabilité, etc.).

Un modèle assis sur DVF+.

Le modèle DV3F conserve la même structure et les mêmes variables que le modèle DVF+.

Les 2 évolutions par rapport à DVF+ sont liées à l'apport des Fichiers fonciers :

- Ajout d'une nouvelle table **acheteur_vendeur** qui regroupe les caractéristiques des propriétaires,
- Ajout de nouvelles variables sur les 3 tables principales **mutation**, **disposition_parcelle** et **local**.

Diagramme du modèle DV3F

Le diagramme du modèle DV3F est accessible ici. Il décrit les champs des 18 tables du modèle et les relations entre elles.

Un enrichissement important des tables principales

Comme pour DVF+, la volonté est de concentrer les informations principales sur les tables principales **mutation**, **disposition_parcelle** et **local** qui comptent chacune plus d'une centaine de variables.

Des variables issues des Fichiers fonciers

En plus des catégories de variables déjà présentes dans DVF+, trois autres catégories existent dans DV3F:

Catégorie	Description	Exemples
Données Fichiers fonciers	Données provenant directement des Fichiers fonciers	ffancst, ffshab, ffsterr, etc.
Données de fiabilité	Indicateurs permettant de s'assurer de la qualité ou de l'exhaustivité d'un rapatriement de données issues des Fichiers fonciers	rapatffloc, fiabilitea, fiabmaxv, etc
Données issues des croisements DVF/FF	Données qui ont pu être déterminées en rapprochant les informations DVF avec celle des Fichiers fonciers	anciennete, segmtab, etc.

Syntaxe des variables

De nouvelles règles syntaxiques viennent également completer celles déjà existantes pour DVF+:

Si la variable commence par...	Alors il s'agit d'un(e)...
ff	donnée directement issue des Fichiers fonciers
l_ff	liste (type array) de données issues des Fichiers fonciers

Si la variable se termine par...	Alors il s'agit d'un(e)...
a	donnée liée à l'acheteur
v	donnée liée au vendeur

Versionnement de DV3F

Chaque nouvelle version de DV3F intégre de nouvelles données sources DVF et/ou des nouveaux millésimes des Fichiers fonciers. Elle s'accompagne dun certain nombre d'évolutions (ajout d'indicateurs, amélioration de méthodologies, etc.). Il n'est donc pas nécessaire de conserver d'anciennes versions de DV3F et il est préférable de travailler avec la dernière version disponible.

Il est important de bien définir les limites temporelles pour le type d'observation que l'on souhaite réaliser. En particulier, lorque l'on effectue des analyses quantitatives (nombre de mutations ou volume de transactions, par exemple), il faut éviter de regarder l'évolution sur les derniers trimestres du fait des temps de latence liés à l'enregistrement des actes de vente et à leurs appartitions dans DVF.

Autre exemple, si l'on travaille sur les jeux d'acteurs dans DV3F, que l'on souhaite une bonne fiabilité et un taux de remplissage optimum sur les acheteurs/vendeurs, il sera préférable de travailler sur la période couverte par les Fichiers fonciers.

Le détail des différentes versions de DV3F produites est décrit dans Note des versions de DV3F