



# Structurer les données

## LA FICHE EN BREF

Dès la mise à disposition des données via le service DVF, une réflexion a été lancée sur leur organisation sous la forme d'un Modèle Conceptuel de Données. Cette représentation permet de structurer les données pour permettre la sécurisation de leur usage, la fiabilisation des traitements et le partage de procédures et d'expériences.

Un important travail collectif, réalisé pendant plusieurs années au sein du Groupe national DVF, a permis d'arriver à un modèle de données stable, fiable et partagé. Des scripts ont été élaborés à partir de ce modèle pour permettre l'importation des données DVF en base de données, facilitant ainsi leur utilisation.

Véritable socle commun, ces scripts sont mis à la disposition de tous les utilisateurs qui peuvent les installer, les adapter puis les utiliser selon leur besoins et usages. Des applications logicielles sont également en cours de développement sur la base de ce socle commun.

### LE MCD

Le Modèle Conceptuel de Données est une représentation de l'organisation des informations et de leurs interrelations.

### LES SCRIPTS

Les scripts permettent un import automatique des informations vers un logiciel de gestion de bases de données.

### LES PREREQUIS

La version actuelle des scripts DVF s'appuie sur des logiciels gratuits, mais requiert une compétence technique.

## 9.1 Qu'est-ce qu'un Modèle Conceptuel de Données ?

Un Modèle Conceptuel de Données (MCD) a pour objectif d'expliciter l'organisation des données qui seront utilisées par le système d'information. Il s'agit d'une représentation des données, facilement compréhensible, permettant de décrire l'information à l'aide d'entités. Le MCD permet de **représenter la structure du système d'information**, du point de vue des données, et définit également les dépendances ou relations entre ces différentes données.

Le modèle résulte de l'analyse conceptuelle des données, laquelle a pour but de chercher des réponses aux questions **que peut-on faire et avec quelles données**.

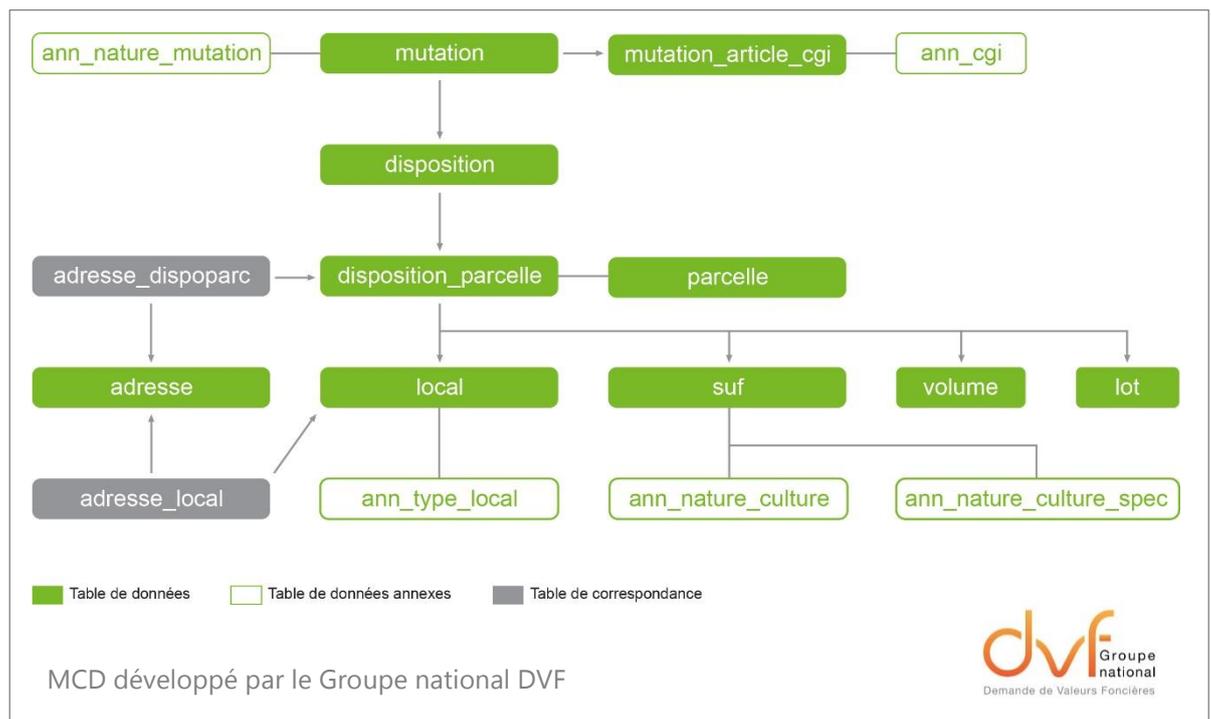
Il a ainsi deux vertus :

- **améliorer la compréhension de la base** de données en maîtrisant l'organisation et donc les possibilités d'analyse
- **faciliter le traitement des données** grâce à la mise en place, sur la base du MCD, de scripts et/ou d'un logiciel

## 9.2 Le MCD développé par le Groupe national DVF

Dès la mise en place de DVF, les utilisateurs ont cherché à structurer les données au sein d'un modèle pour mieux les comprendre et mieux les exploiter. Les travaux autour de la création d'un modèle s'inscrivent dans un processus qui a débuté en 2011 à l'initiative du Groupe national DVF. Grâce à un travail collaboratif, un Modèle Conceptuel de Données (MCD) pour l'import des données natives de DVF a été produit et partagé collégialement par le groupe. Ce modèle est le résultat d'un **partage d'une vision commune**.

Le MCD proposé par le Groupe national a pour vocation de devenir le **référentiel commun** à tous les utilisateurs DVF. Face à la multiplication des modèles de données développés par chaque structure utilisatrice, il semblait opportun de se diriger vers une harmonisation en suggérant l'utilisation d'un modèle commun permettant des usages variés. Le schéma présenté ci-dessous est une synthèse du modèle à jour.





## 9.5 Le processus d'intégration des scripts

### 9.5.1 Mettre en œuvre l'import des données en 3 étapes



## 9.5.2 Précautions d'usages

**ETAPE 1 :** Le script « 00\_creation\_base\_vide.sql » n'est à exécuter que la première fois pour créer la base vide. En cas de réexécution, la base DVF serait supprimée sans message d'avertissement.

**ETAPE 2 :** Trois lignes (bien indiquées dans le script) sont à modifier dans ce fichier afin de spécifier les noms et chemins des fichiers annexes. Les chemins d'accès aux données brutes ne doivent pas comporter de caractères spéciaux, espaces, signes de ponctuation, etc.

**ETAPE 3 :** Une ligne (bien indiquée dans le script) est à modifier dans ce fichier afin de spécifier le nom et chemin du fichier principal. Il ne faut importer qu'un seul fichier à la fois.

### A NOTER

Sur un ordinateur récent, il faut compter moins d'une minute pour l'import d'un fichier départemental de DVF avec les scripts.

**3 étapes à effectuer avec précaution**

## 9.6 Aboutissement de l'intégration des scripts

	id_mutation [PK] integer	code_service_ch character varying(10)	reference_document character varying(10)	date_mutation date	annee_mutal integer	mois_mutati integer	id_nature_mutation integer	lib_nature_mutation character varying(254)
1	1	5914P03	2014P01980	2014-03-20	2014	3	3	Vente
2	2	5914P02	2014P02365	2014-01-29	2014	1	3	Vente
3	3	6204P05	2014P00690	2014-02-13	2014	2	3	Vente
4	4	5924P01	2014P02325	2014-05-14	2014	5	3	Vente
5	5	5914P04	2014P02660	2014-06-03	2014	6	3	Vente
6	6	6204P01	2014P02974	2014-05-12	2014	5	3	Vente
7	7	5914P02	2014P05028	2014-04-17	2014	4	3	Vente
8	8	5914P02	2014P02815	2014-02-25	2014	2	3	Vente
9	9	5914P02	2014P03265	2014-02-18	2014	2	3	Vente
10	10	5914P06	2014P02704	2014-06-19	2014	6	3	Vente
11	11	5914P02	2014P05717	2014-05-16	2014	5	4	Vente en l'état futur d'achève
12	12	5914P03	2014P02989	2014-05-27	2014	5	3	Vente
13	13	6204P05	2014P01182	2014-02-28	2014	2	3	Vente
14	14	5914P02	2014P04683	2014-04-02	2014	4	3	Vente
15	15	5914P02	2014P05654	2014-05-06	2014	5	3	Vente
16	16	6204P02	2014P02144	2014-02-27	2014	2	3	Vente
17	17	6204P04	2014P00721	2014-01-23	2014	1	3	Vente
18	18	5914P03	2014P03725	2014-06-25	2014	6	3	Vente
19	19	5914P01	2014P01003	2014-01-06	2014	1	4	Vente en l'état futur d'achève
20	20	5924P03	2014P01935	2014-02-27	2014	2	3	Vente
21	21	6204P06	2014P00913	2014-02-07	2014	2	3	Vente
22	22	5914P06	2014P01391	2014-03-17	2014	3	3	Vente
23	23	6204P06	2014P00529	2014-01-24	2014	1	3	Vente
24	24	6204P03	2014P02634	2014-05-16	2014	5	3	Vente
25	25	6204P01	2014P03148	2014-06-17	2014	6	3	Vente

L'exécution des scripts SQL permet le chargement des données DVF dans plusieurs tables de manière structurée, selon un modèle de données issu du MCD validé par le GnDVF (ci-dessus, un extrait de la table Mutation).

Cette table a l'avantage de présenter une seule ligne par mutation et les mutations sont caractérisées, ce qui permet notamment d'effectuer aisément une typologie des biens concernés (voir fiche 8 *Typologie des biens*).

**Une base de données optimisée grâce à l'intégration des scripts**

## 9.7 Des scripts SQL à l'application ImportDVF+

Une application pour simplifier l'usage des scripts



The screenshot shows the 'AppDVF' web application interface. At the top, there is a navigation bar with the 'AppDVF' logo and a hamburger menu icon. Below this, the main heading is 'ImportDVF+'. The central part of the page is a form titled 'Paramétrage de la base de données'. The form contains several fields: 'Hôte' (localhost), 'Base de données' (dvh), 'Utilisateur' (user), 'Mot de Passe' (masked with dots), 'Port' (5432), 'Dossier données DVF' (C:\AppDVF\appdvf\_vbeta-3), 'Géométries (1)' (checkbox), 'Proxy (1)' (http://proxy.example.fr - laissez vide dans un premier temps), and 'Effacer schemas (2)' (checkbox). Below the form, there are two footnotes: (1) - La fonctionnalité permettant de récupérer les géométries depuis cadastre.api.gouv.fr est indisponible, le service n'étant actuellement plus actif. (2) - En cochant la case, tous les schémas DVF+ départementaux existants seront supprimés de la base de données. Dans le cas inverse, ce sont uniquement les schémas départementaux liés aux données à intégrer qui seront écrasés, les autres seront conservés. At the bottom of the page, there is a footer with 'Crédits - Contact' and 'AppDVF - Cerema - BY-NC-SA'.

Afin de faciliter l'intégration des données brutes DVF dans une base de données PostgreSQL, le Cerema a développé sous licence libre l'application Import DVF+. Cet outil s'appuie sur les scripts SQL (présentés au paragraphe 9.3). Il permet donc d'importer les données en respectant le modèle de données validé par le Groupe national DVF.

Comme pour l'utilisation des scripts SQL, il est nécessaire de disposer du logiciel adéquat (PostgreSQL) et d'avoir la dernière version de l'application ImportDVF+.

ImportDVF+ s'intègre désormais dans une interface plus large appelée AppDVF (voir encadré).

La dernière version de l'application AppDVF est disponible sur le site :

<http://www.geoinformations.developpement-durable.gouv.fr/appdvf-un-outil-libre-pour-l-exploitation-de-dv3f-r1041.html>

### LE SAVIEZ-VOUS ?



L'application ImportDVF+ a été intégrée dans une application plus riche, également sous licence libre, appelée AppDVF.

AppDVF est une interface libre développée depuis 2017 par le Cerema et la DGALN. Elle réunit désormais toutes les applications développées par le Cerema et la DGALN pour faciliter l'exploitation des données DVF (voir également 9.8.2).

## 9.8 Les pistes d'amélioration et d'enrichissement

Le MCD, les scripts et l'application ImportDVF+ sont des outils efficaces pour traiter la base de données DVF. Néanmoins, ils sont perfectibles. Plusieurs réflexions et travaux sont en cours.

### 9.8.1 Evolution potentielle du MCD

L'appréhension de DVF s'améliore au sein de la communauté des utilisateurs et les usages se multiplient. Le MCD est donc amené à évoluer afin de prendre en compte le nouveau niveau de connaissance de la base de données et les nouveaux traitements envisagés.

La mise en place de nouveaux outils comme les scripts ou les logiciels questionne aussi le MCD et pousse à envisager sa mise à jour.

**Un MCD  
mis à  
jour ?**

### 9.8.2 Amélioration prévue des scripts et applications existants

A la suite de l'envoi par les utilisateurs de retours d'expérience, le Cerema reprend au fur et à mesure les scripts et les applications pour les améliorer. Ces correctifs prennent en compte les erreurs et difficultés rencontrées par les utilisateurs dans l'intégration et dans le fonctionnement des scripts.

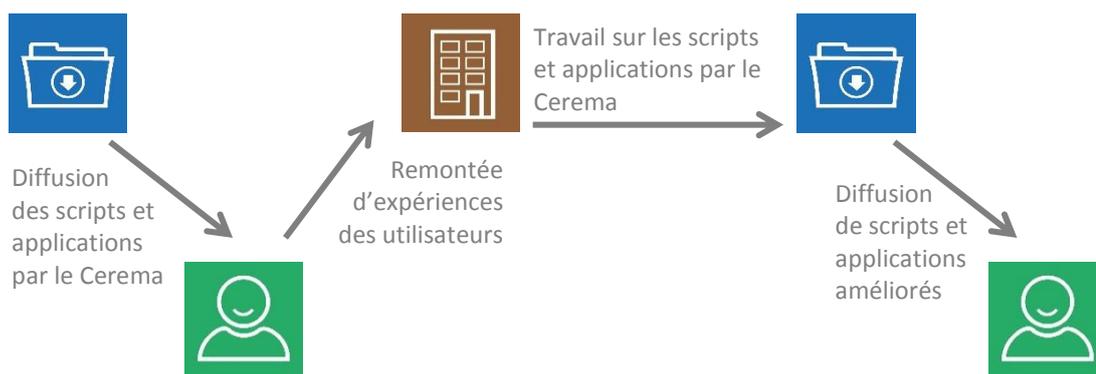
Le Cerema va ainsi livrer prochainement une version stabilisée de l'application AppDVF. Le concept repose sur l'intégration de nombreux indicateurs directement dans la table des mutations pour permettre aux utilisateurs

d'accéder plus facilement aux données contenues dans DVF sans connaissance du langage SQL. Il sera possible alors d'utiliser la table directement via un tableur. Le fichier fera apparaître une mutation par ligne.

Une version de test (bêta) de l'application AppDVF est en ligne depuis septembre 2017 à l'adresse suivante :

<http://www.geoinformations.developpement-durable.gouv.fr/appdvf-un-outil-libre-pour-l-exploitation-de-dv3f-r1041.html>

**Des scripts  
et  
applications  
améliorés  
grâce aux  
retours des  
utilisateurs**



### LE SAVIEZ-VOUS ?

L'EPF Bretagne a développé un complément aux scripts du Cerema pour proposer l'enrichissement de DVF. C'est une table qui permet l'insertion d'autres sources de données afin de les croiser avec DVF.



### 9.8.3 Poursuite du développement de logiciels sur le socle commun

Si aujourd'hui le développement d'un logiciel n'est pas la priorité du Groupe national DVF, le Modèle Conceptuel de Données constitue déjà une première étape vers la création de logiciels comme le montre le développement des applications ImportDVF+ et AppDVF.

Le modèle national et les scripts permettent aux éditeurs de logiciels de bénéficier des travaux de recherches menés par le Groupe national DVF

(GnDVF). Ceux-ci sont un gage de stabilité, de fiabilité et de pérennité pour eux. Ils peuvent ainsi s'appuyer sur l'expertise du GnDVF pour développer des outils.

Quelques utilisateurs, comme Quélleville? ou l'EPF Bretagne, ont travaillé à l'élaboration d'un logiciel. L'objectif est de consulter et de valoriser les données DVF dans un environnement métier sans avoir à redévelopper une application.

### 9.8.4 Faciliter les enrichissements de DVF par d'autres données

La structuration des données DVF grâce au Modèle Conceptuel de Données et leur intégration dans une base de données facilite leur exploitation. Notamment, cela permet d'enrichir les données DVF par croisement avec d'autres sources de données : couches d'information géographique, fichiers fonciers du cadastre, zonages d'urbanisme, etc.

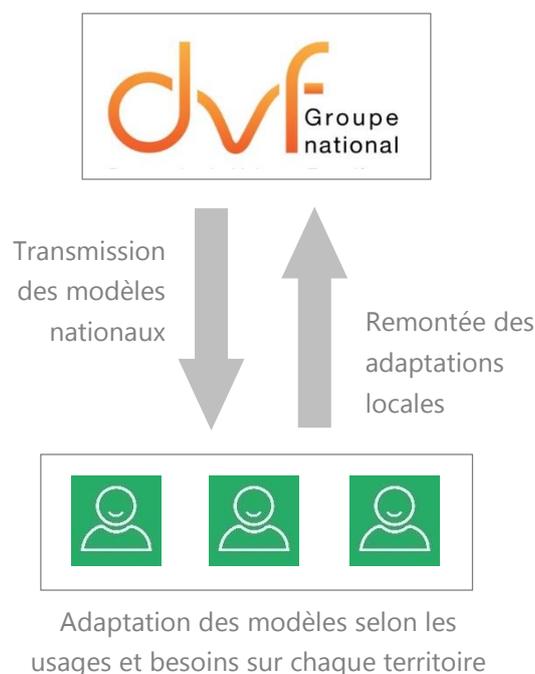
Plusieurs expériences vont dans ce sens. Par exemple, le Cerema a utilisé sa base des

Fichiers Fonciers pour enrichir les données DVF dans une autre base appelée DV3F (DVF+Fichiers Fonciers).

Le croisement avec l'information de la géographie des parcelles permet ainsi la représentation cartographique des mutations enregistrées dans DVF. Cet usage particulier est détaillé dans la *fiche 10 : La spatialisation des données DVF* de ce guide.

### 9.8.5 Le rôle du Groupe national

Le Groupe national DVF joue un rôle central de coordination de l'ensemble de ces démarches. Il est important de mutualiser les travaux et les axes de recherche. Chaque piste d'amélioration envisagée par une structure doit pouvoir bénéficier à l'ensemble des utilisateurs de DVF. Le Groupe national DVF veille ainsi à ce que les pistes d'amélioration menées par les utilisateurs soient cohérentes et harmonisées dans le cadre d'une stratégie nationale visant à faciliter toujours plus l'exploitation des données DVF tout en maintenant un référentiel commun. Le Groupe national s'assure ainsi que les travaux ne soient pas dédoublés mais au contraire mutualisés.



**Un socle commun pour les logiciels**

**Un socle commun pour les enrichissements**

**Le Groupe national coordonne les démarches : mutualiser pour mieux utiliser**