

SIG BDD

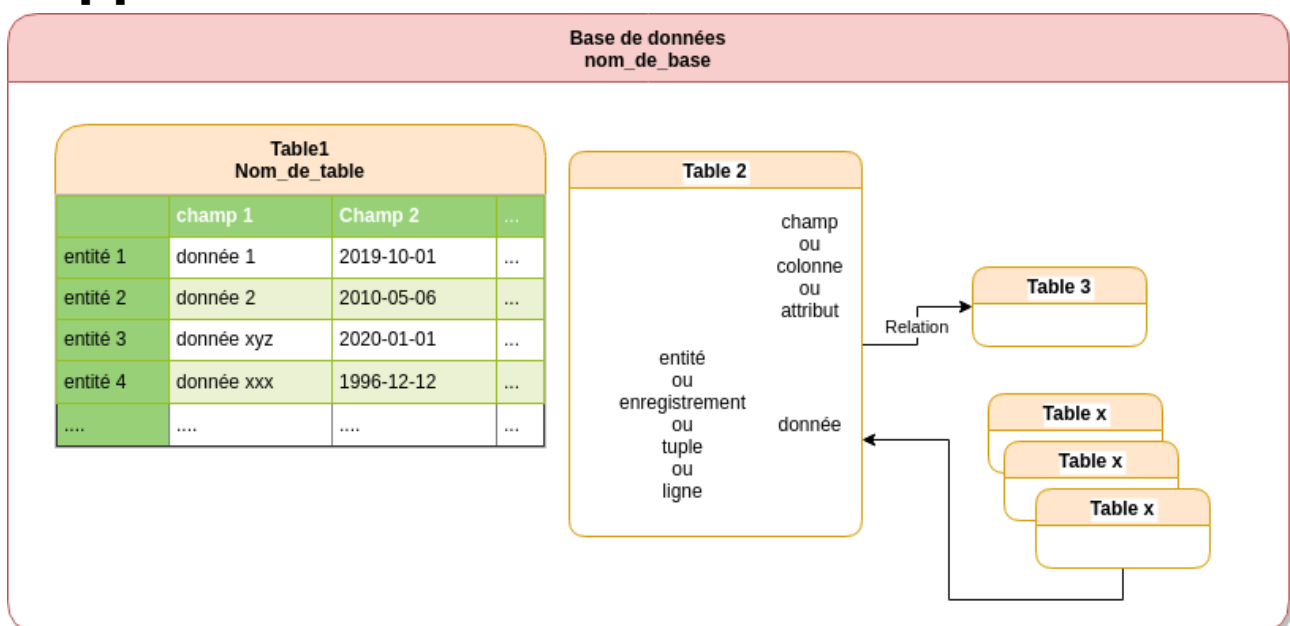
CONCEVOIR UNE BASE DE DONNÉES

Une base de données (ou la partie attributs d'une couche SIG) correctement conçue vous permet d'accéder aux informations à jour et précises. Une conception correcte est essentielle pour atteindre vos objectifs d'utilisation d'une base de données.

Pour cela, quelques étapes et principes sont indispensables.

Cette fiche ne représente que quelques principes, elle n'aborde pas les thématiques de modélisation avancée, ne cite aucune méthode de conceptualisation, et n'est pas une méthode de modélisation. Ce sont juste quelques conseils.

Rappel sur la structure d'une base de donnée.



Une couche SIG ne contient qu'une seule table !!

Objectif de la base

La première question est forcément de se demander à quoi va servir votre base ou couche SIG.

Pour cela,

- définir qui va l'utiliser,
- comment ou dans quel contexte,
- qui va la remplir ou la mettre à jour...

Il s'agit de **rédiger un cahier des charges** (court, 4 ou 5 lignes) qui permettra de bien cibler les attendus aux moments de prendre des décisions dans la conception.

Analyse des besoins

Une première étape est de commencer la rédaction par : « Je veux pouvoir... » (faire, afficher, analyser, sortir...) auprès des utilisateurs.

On s'appuie également sur le recueil de documents (tableur, cahier des charges, autres bases, interviews, factures, listings...). Ces documents peuvent déjà exister ou doivent être conçus lors de cette phase.

Avec ces éléments,

- **rédiger ou imaginer les rapports et les sorties nécessaires.**
- **imaginez-le ou les formulaires pour enregistrer** les informations nécessaires,

Cela ne sera pas parfait du premier coup, mais il est important de tout référencer et **de rassembler un maximum d'avis, de poser des questions aux utilisateurs.**

Exemple : Si votre but est d'envoyer un email à Monsieur Machin, il faut que vous ayez la civilité (monsieur), le nom (machin), le mail...

De la qualité de cette étape dépend la pertinence de la base de données par rapport aux usages.

Plus d'un projet sur deux est un échec, car les besoins des utilisateurs sont mal pris en compte.

Conception de la base

Premièrement, utilisant l'analyse faite avant, **transformer les éléments d'information en table et colonnes.**

Exemple d'un premier jet : envoyer un mail à M. Machin

table : adresse_envoi						
Nom des colonnes	civilité	Nom	adresse	mail	age	...
Exemple de données	M.	MACHIN	6 route de machin 29000 MACHINVILLE	machin@machintech.com	40	

Une fois que vous avez déterminé le premier jeu de colonnes pour chaque table, vous devez affiner davantage les colonnes en appliquant les règles d'or suivantes :

– Stocker les informations dans leur plus petite partie logique.

Par exemple, il est préférable d'enregistrer le nom d'un client dans deux colonnes distinctes : prénom et nom de famille, afin que vous puissiez trier, rechercher. De même, l'adresse se compose de cinq éléments distincts : adresse, ville, état, code postal et pays/région, et il est donc judicieux de les stocker dans des colonnes distinctes.

– Ne pas stocker de données calculées.

Elles gâchent de l'espace et nécessitent des mises à jour très régulières. De plus, généralement est fait le calcul avant d'extraire la donnée pour quelle soit à jour, donc

autant faire le calcul directement... Cela inclut de stocker une date de naissance plutôt qu'un âge par exemple.

– Pas de redondances.

Vous devez éviter la répétition d'information entre les colonnes et entre les tables.

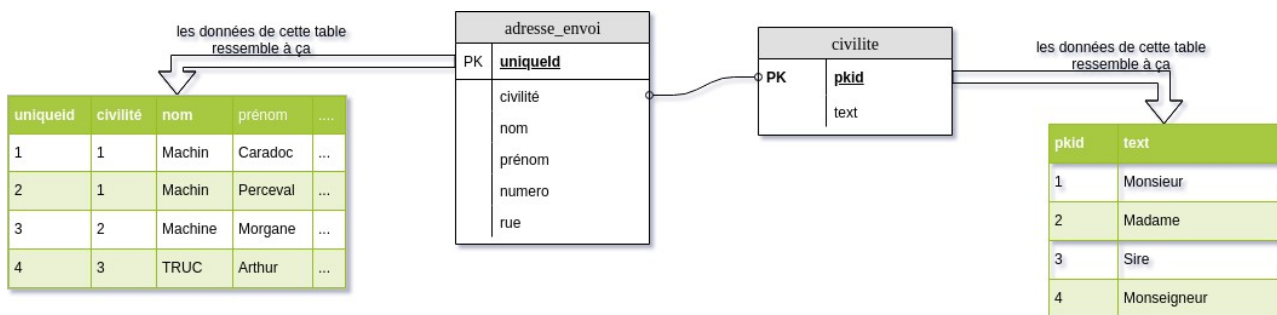
– Penser à la propreté des saisies

Dans l'exemple, la civilité peut être saisie suivant l'opérateur de nombreuses façons : M., M, Monsieur, MONSIEUR, mnosiuer.....

Pour cela, 2 solutions :

- soit limiter au niveau de l'application de saisie : liste de valeurs dans QGIS, par exemple
- introduire une autre table gérant ces données et créée une relation qui permettra d'encadrer la saisie de cette colonne

Je n'aborde pas ici le thème des bases de données relationnelles et des différentes relations (cf. un exemple sur l'illustration)



Dans le choix de la propreté, il faut penser au **type de données des colonnes**, cela évite par exemple de saisir des lettres dans une date... :

Alphabétique : lorsque la donnée est uniquement composée de caractères alphabétiques (de "A" à "Z" et de "a" à "z") (peu courant)

AlphaNumérique : lorsque la donnée peut être composée à la fois de caractères alphabétiques et numériques (donnée type TEXT)

Numérique : lorsque la donnée est composée uniquement de nombres (**entiers ou réels**)

Date : lorsque la donnée est une date (au format AAAA-MM-JJ) avec ou sans heure

Booléen : Vrai ou Faux

– Inclure ou choisir un identifiant unique

À FAIRE SYSTÉMATIQUEMENT MÊME POUR UNE SEULE TABLE. On ne sait jamais ! La plupart des bases de données n'autoriseront pas de table sans clé primaire

Pour effectuer une mise en relation ou éviter de confondre des entités, il faut créer ou trouver un **identifiant unique** qui permettra de faire les relations mais aussi les jointures.

C'est ce que l'on appelle la clé primaire (PK ou primary key en anglais) : **elle doit être unique (pas de doublon possible) et jamais NULL (c'est-à-dire vide).**

La plupart du temps, il s'agit d'une numérotation fait automatiquement par le logiciel de base de donnée, mais vous pouvez utiliser une colonne existante (exemple : le code INSEE dans une table de COMMUNES).

—

Il en est de même d'une base de donnée que d'une couche SIG, la bonne préparation et la connaissance des usages permet d'éviter des erreurs. Cependant, aucune base ou couche n'est parfaite du premier coup, n'hésitez pas à rajouter des tables/colonnes ou à en retirer : mais **ATTENTION À NE PAS PERDRE DE DONNÉES, cela doit être la priorité.**