

# OIM SIG : TP2

## Analyse de données

---

### 1. Déroulement

Nous allons dans un premier temps manipuler des données afin de comprendre le fonctionnement du langage SQL, ensuite, vous mettrez en application les différentes instructions vues dans le cadre des données sur la Guadeloupe.

Récupérez le fichier de données à l'adresse suivante (vous aurez également besoin des données du TP1).

<http://calamar.univ-ag.fr/uag/ufrsen/coursenligne/egrandch/sig/ls4/DataTP2LS4.zip>

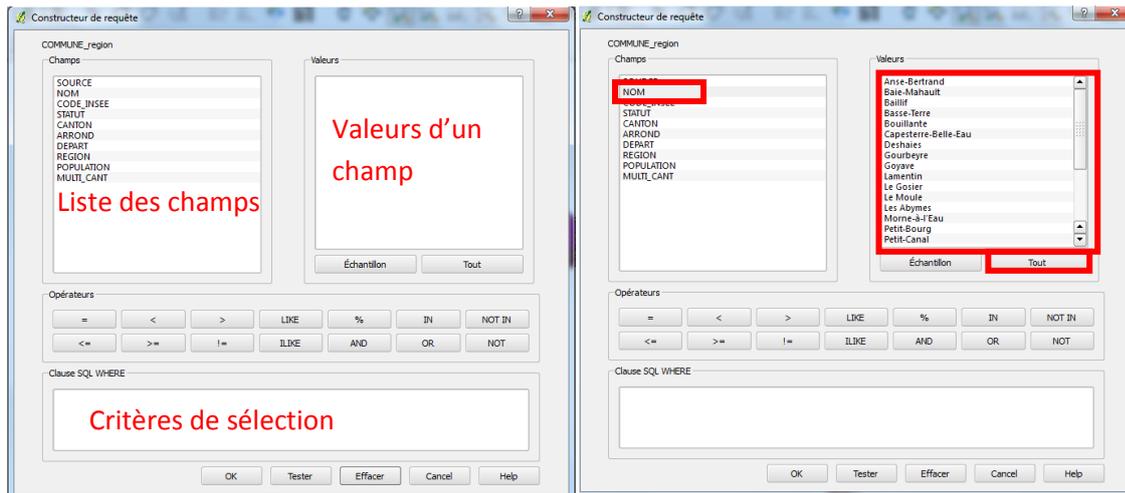
Décompresser le fichier zip. Ouvrir les différents shapes (.shp) sous QGIS.

### 2. Remarques préliminaires

Il existe plusieurs façons d'analyser des données sous *QGIS* en fonction de la complexité des analyses souhaitées et des outils dont on dispose. Ces actions sont liées au langage **SQL**, qui est un langage permettant d'accéder aux données d'une base de donnée, et ses extensions pour les données spatiales. Les fonctionnalités les plus simples sont disponibles dans la version de base de *QGIS* et permettent de sélectionner de l'information en se basant soit sur la valeur d'un ou plusieurs champs d'une couche d'information soit sur des propriétés spatiales pour sélectionner par exemple un objet à l'intérieur d'un autre ou ayant une intersection avec un autre. Pour pouvoir disposer de fonctionnalités avancées comme la sélection d'objets en fonction de leurs caractéristiques géométriques (surface, périmètre, etc.) il faut utiliser une base de données spatiale plus complète que les couches d'informations contenant les objets. Pour cela plusieurs solutions sont possibles et nécessitent l'installation d'extensions de *QGIS*. Nous citerons ici deux extensions parmi les plus utilisées : le couple *PostgreSQL/PostGIS* et le couple *SQLite/Spatialite*. Le premier est très complet et supporte des bases de données en réseau avec un modèle client/serveur, il requiert l'installation du *SGBD PostgreSQL*. Le second est plus simple à installer et est dédié à une utilisation sur un poste unique de la base de données. Nous verrons dans le cadre de ce TP la seconde solution.

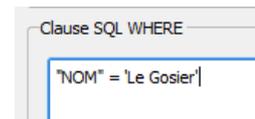
### 3. Sélection de données à partir des valeurs des champs

Pour sélectionner les données d'une couche en filtrant sur la valeur d'un ou plusieurs champs. Faites un clic droit sur la couche et sélectionnez l'entrée *Requêtes*. La fenêtre suivante s'ouvre.

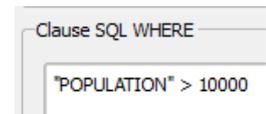


En haut à gauche se trouve la **liste des champs** de la couche (ici la couche *Commune*). En haut à droite peuvent être affichées les **valeurs d'un champ**. Pour cela **sélectionnez un champ** et cliquez sur **Echantillon** (pour avoir un aperçu des valeurs du champ) ou sur **Tout** pour visualiser l'ensemble des valeurs (Figure de droite). Dans la partie du bas, on peut rentrer des **critères de sélection** pour filtrer les données.

Par exemple, si on souhaite sélectionner la commune *Le Gosier* on entrera .



Si on souhaite sélectionner les communes ayant plus de 10000 habitants, on entrera.



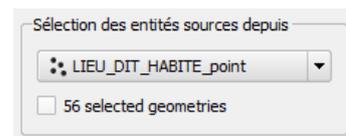
**Remarque** : Il est conseillé d'utiliser l'interface pour saisir les noms de champs, les valeurs et les opérateurs. Pour cela il suffit de cliquer ou double cliquer en fonction du cas sur les différents éléments de l'interface. Pour une aide sur la formulation des requêtes, reportez vous à [QgsQueryBuilder](#)

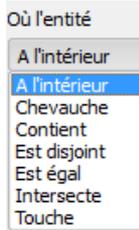
## 4. Sélection de données à partir de l'information spatiale

### 4.1 Entrée **Requête spatiale** du menu **Vecteur**.

**Remarque** : Si l'entrée n'existe pas dans le menu vecteur c'est que l'extension n'est pas activée. Dans ce cas, allez dans le menu **extension/gestionnaire d'extensions** et cochez **Extension Requête Spatiale**.

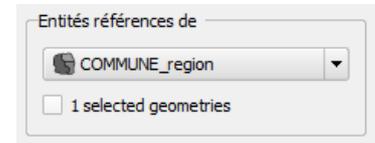
On précise alors à partir de la fenêtre de saisie la couche dans laquelle va se faire la sélection (**Sélection des entités sources depuis**). Ici par exemple la couche sur les *Lieux Dits*.





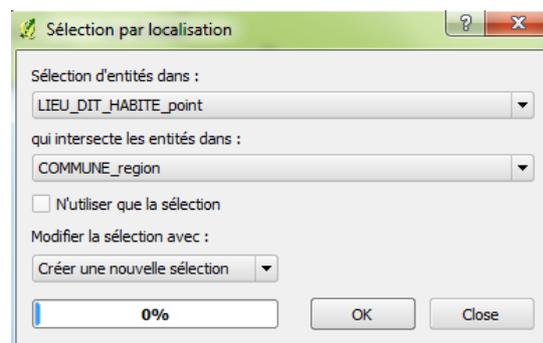
On sélectionne ensuite le critère spatial (*Où l'entité*): à l'intérieur, Chevauche, Contient, Est disjoint, Est égal, Intersecte ou Touche. Cette liste des critères dépend du type, point, ligne ou polygone, des couches utilisées.

On sélectionne ensuite la couche devant servir pour la sélection (*Entités références de*).



#### 4.2. l'entrée *Sélection par localisation* du sous menu *Outils de Recherche* du menu *Vecteur*

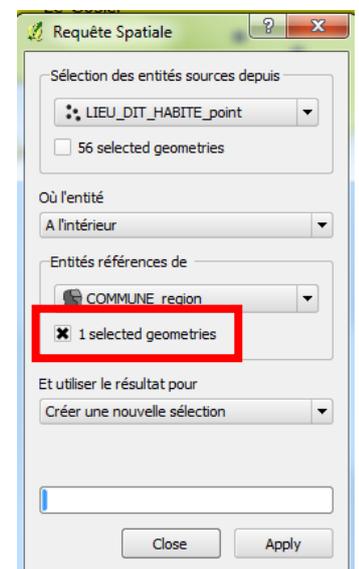
Cet autre moyen est plus limité dans les possibilités.



### 5. Combinaison de sélections

On peut combiner les différentes sélections précédentes pour par exemple sélectionner les lieux dits de la commune du Gosier. On procède pour cela de la manière suivante :

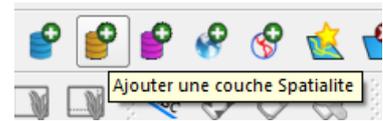
1. Sélection de la commune du Gosier à l'aide d'une requête sur les attributs.
2. Sélection des lieux dits à l'intérieur de la sélection précédente à l'aide d'une requête spatiale en cochant la sélection.



### 6. Requêtes SQL plus complexes avec Spatialite

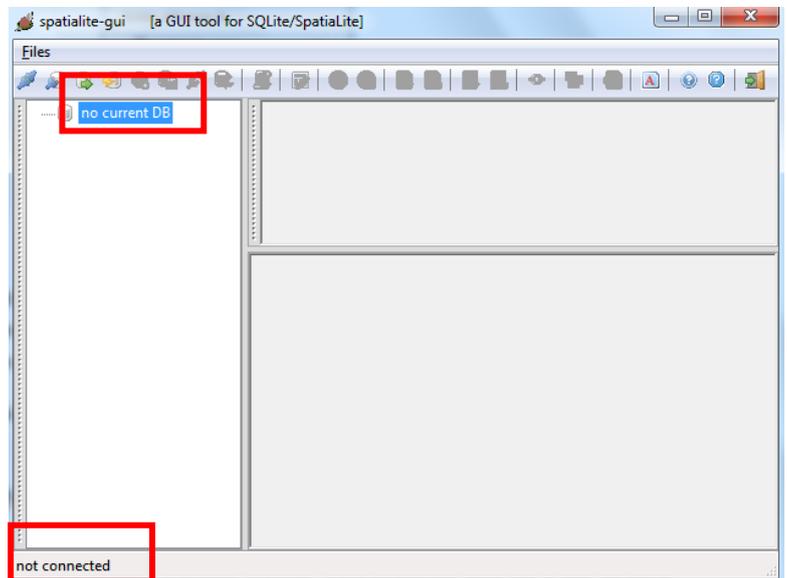
En utilisant l'extension *Spatialite* qui se trouve dans le menu *Base de données*, on peut également sélectionner des enregistrements en fonction de leurs caractéristiques géométriques (surface, périmètre, etc.).

L'utilisation de *Spatialite* permet d'ajouter une couche *Spatialite* (à partir d'une base de donnée créée comme précédemment) à l'environnement *Qgis* de la même manière que l'on ajoute une couche *vecteur* ou *raster*.



### 6.1 Préparation de la base de donnée

Avant d'utiliser *Spatialite*, il faut transformer les couches d'information en véritables bases de données en utilisant l'application *spatialite\_gui* (qui se trouve dans le fichier de données *DataTP2LS4.zip*). Lancez l'application et la fenêtre suivante s'ouvrira, vous indiquant qu'il n'y a pas de base de données actuellement créée.



Créez une nouvelle base de données et nommez la *maBase.sqlite*



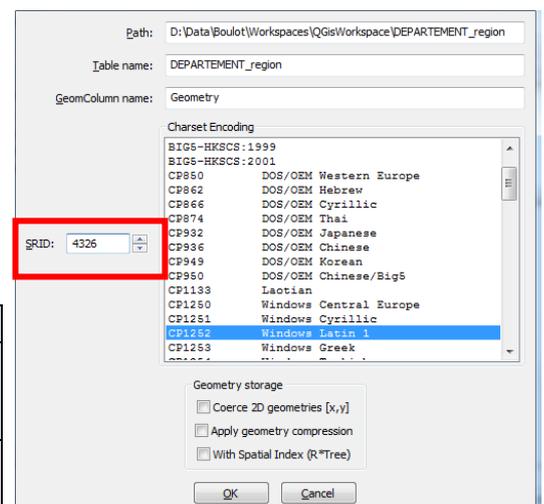
Créez ensuite un *shapefile virtuel* et sélectionnez le fichier *DEPARTEMENT.shp*.



Saisir la valeur 4326 pour le SRID. Le SRID (*Spatial Reference System Identifier*) est un code international permettant d'identifier le référentiel et le système de projection utilisé, il existe des SRID 2D et 3D.

Voici quelques SRID 2D utiles (liste complète disponible sur <http://spatialreference.org>)

SRID	Nom	Système
4326	Standard mondial GPS	WGS84, [long-lat], unités en degré (distance) et degrés carrés (surface)
2970 ou 32620	Guadeloupe/Martinique	WGS84, UTM 20 N
2972	Guyane	RGFG95, UTM 22 N
27581	France	NTF France I



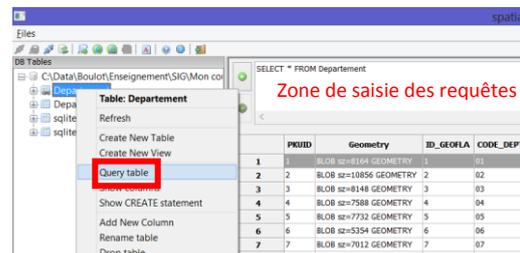
## 6.2 Exécution d'une requête SQL dans spatialite-gui

Nous pouvons maintenant créer nos requêtes soit en restant dans cette interface soit en lançant *QSpatialite* depuis le menu *Base de Donnée/Spatialite* de *QGIS*. La première solution permet de gérer de manière plus complète la base de donnée.

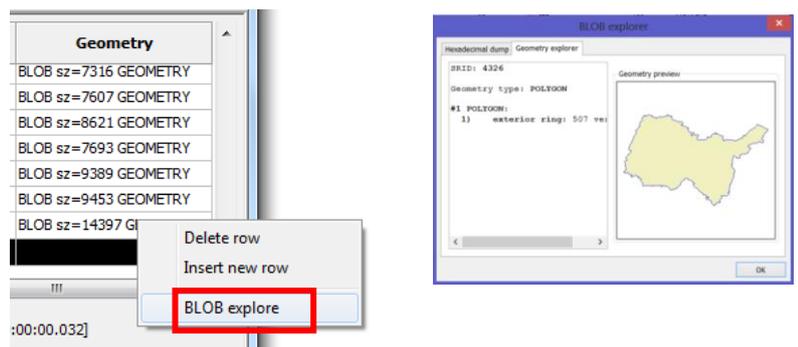
Pour exécuter une requête SQL il faut saisir la requête dans la **zone de saisie** puis cliquer sur



Pour afficher le contenu de la table : **cliquez droit** sur la table *département*, puis **Query table**



Pour afficher l'information géographique : **cliquez droit** sur le champ *Geometry* d'une des lignes et en sélectionnant **BLOB explore**.

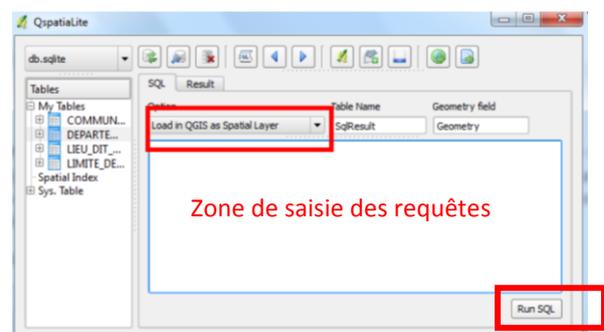


## 6.3 Exécution d'une requête SQL dans QGIS

Lancez *QSpatialite* qui se trouve dans le menu *Base de Donnée/Spatialite*

Une fois la base de donnée créée et les tables ajoutées sous *spatialite-gui*, il est préférable d'utiliser l'outil *spatialite* intégré dans *QGIS* car il permet l'affichage du résultat des sélections directement dans *QGIS* (en sélectionnant dans **option Load in QGIS as spatial Layer**).

Pour exécuter une requête SQL il suffit de la saisir dans la **zone de saisie** puis de cliquer sur **Run SQL**



## 7. Exemples de requêtes SQL

Les requêtes SQL peuvent être construites de multiples manières, le lien suivant vous permettra d'apprendre comment construire des requêtes si vous le souhaitez.

<http://www.gaia-gis.it/spatialite-2.4.0-4/spatialite-cookbook-fr/index.html#common>

La liste des fonctions SQL est disponible ici

<http://www.gaia-gis.it/spatialite-2.4.0-4/spatialite-sql-2.4-4.html>

Voici quelques exemples de requêtes possibles sur les tables qui vous sont fournies

Les valeurs renvoyées par les différentes fonctions sont dans les unités de la couche (par exemple pour le SRID 4326, ce sont des degrés pour les distances et des degrés carrés pour les surfaces).

Si on veut avoir des résultats dans d'autres unités il faut convertir au préalable les géométries en utilisant la fonction `Transform("MaCouche".'Geometry',NouvelSRID)` en spécifiant le SRID correspondant aux unités que l'on souhaite avoir.

Requêtes sur la table « DEPARTEMENT »

- Liste des départements de plus de 8500 km<sup>2</sup>: `SELECT * FROM "DEPARTEMENT" where Area(Transform("DEPARTEMENT".'Geometry',27581))/1000000 > 8500`  
(La fonction Transform permet de projeter les données dans le SRID 27581 correspondant à la France avec des unités exprimées en mètre).
- Nombre de départements de plus de 8500 km<sup>2</sup>: `SELECT count(*) FROM "DEPARTEMENT" where Area(Transform("DEPARTEMENT".'Geometry',27581))/1000000 > 8500`  
(La fonction Transform permet de projeter les données dans le SRID 27581 correspondant à la France avec des unités exprimées en mètre).
- Superficie totale de la région Centre: `SELECT sum(Area(Transform("DEPARTEMENT".'Geometry',27581))/1000000) FROM "DEPARTEMENT" where NOM_REGION like "Centre"`
- Nombre de départements dont le nom commence par "N": `SELECT count(*) FROM "DEPARTEMENT" where NOM_DEPT like "N%"`
- Nombre de départements dont le nom contient la chaîne "oi": `SELECT count(*) FROM "DEPARTEMENT" where NOM_DEPT like "%oi%"`
- Périmètre total des départements dont le nom commence par "O": `SELECT sum(length("DEPARTEMENT".'Geometry')) FROM "DEPARTEMENT" where NOM_DEPT like "O%"`

Requêtes sur la table « LIMITE\_DEPARTEMENT »

- Longueur totale par type de frontière: `SELECT NATURE, sum(length("LIMITE_DEPARTEMENT_polyline".'Geometry')) FROM "LIMITE_DEPARTEMENT_polyline" Group by NATURE` (ici on regroupe les lignes par NATURE, puis on fait le calcul `sum(len(obj))` pour chaque groupe).

## 8. Les Jointures

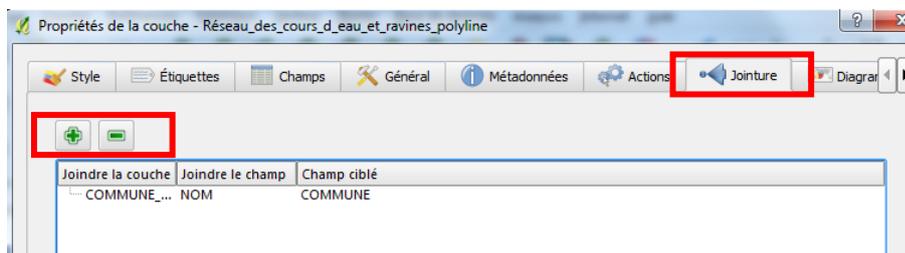
La jointure est l'association de deux ou plusieurs couches d'information. Il s'agit donc d'ajouter les attributs de la seconde couche aux attributs de la première. Il faut pour cela définir un critère permettant d'affecter les valeurs des lignes de la seconde couche aux valeurs des lignes de la première. Par exemple, si on veut associer les champs de la couche *Commune* (nom de la commune, population, ...) à la couche sur les *Lieux Dits*, il faut qu'un lieu dit situé sur la commune du *Gosier* se

trouve associé aux valeurs des champs de cette commune. Pour cela, on peut procéder de deux manières différentes en fonction de l'information dont on dispose.

### 8.1 Jointure sémantique

La jointure sémantique consiste à associer deux couches à partir d'un champ commun. Pour réaliser la jointure sémantique on utilise l'onglet *Jointure* de la fenêtre des propriétés de la couche à laquelle on veut joindre une autre couche.

Pour joindre la couche *commune* à la couche sur les *réseaux de cours d'eau* on utilisera le champ commun contenant le nom de la commune (il s'appelle **Nom** dans la couche **Commune** et **Commune** dans la couche **réseaux de cours d'eau**). L'affichage du tableau des attributs permettra de visualiser l'ensemble des attributs.

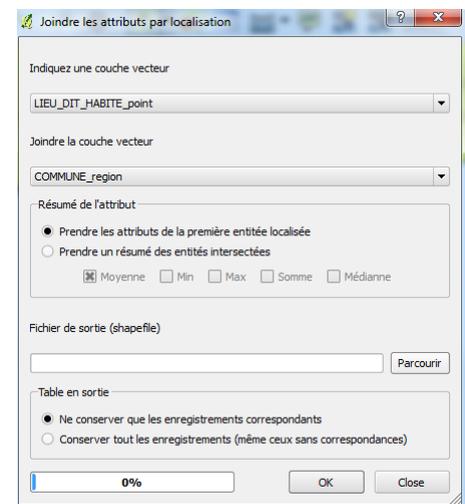


### 8.2 Jointure spatiale

La jointure spatiale consiste à associer deux couches à partir de l'information spatiale. Par exemple, associer à un enregistrement (une ligne) de la couche *Lieux dit* la valeur des champs de la couche *Commune* de l'enregistrement contenant le lieu dit.

Pour cela on utilise l'outil *joindre les attributs par localisation* dans le menu *outils de gestion de données* du menu *vecteur*.

En fonction des cas il arrive que plusieurs lignes de la seconde couche puissent être associées à une ligne de la première couche. Dans ce cas, on peut soit associer la première couche trouvée soit calculer une valeur (moyenne, minimum, maximum, somme, médiane) qui résume l'ensemble des champs (numériques) trouvés dans les différentes lignes.



Le résultat est sauvegardé dans une nouvelle couche et on peut décider de conserver uniquement les lignes de la première couche ayant une correspondance avec au moins une ligne de la seconde couche ou alors conserver toutes les lignes (il y a aura dans ce cas des champs vides pour ces lignes).

### 8.3 Jointure spatiale à l'aide d'une requête SQL

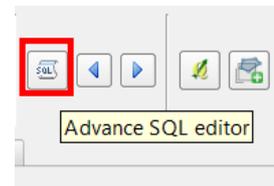
On peut également faire des jointures spatiales avec *spatialite* de manière plus poussée. Par exemple pour avoir la liste des lieux dits de la commune du *Gosier* on écrira dans la fenêtre *spatialite* la requête suivante

```
select * from COMMUNE, LIEU_DIT_HABITE where
Contains (COMMUNE.Geometry, LIEU_DIT_HABITE.Geometry) and COMMUNE.Nom like
"%Gosier%"
```

Un certain nombre de fonctions permettant de faire les jointures sont disponibles telles que *Equals*, *Disjoint*, *Touches*, *Within*, *Overlaps*, *Crosses*, *Intersects*, *Contains*, *Covers*, *CoveredBy*, *Relate*, etc. (<http://www.gaia-gis.it/gaia-sins/spatialite-sql-3.0.0.html>)

## 9. Requêtes à écrire

A partir des données du TP1 sur la Guadeloupe écrivez les requêtes suivantes dans l'outil *spatialite* de QGIS (pensez à noter les requêtes une fois testées et validées pour ne pas les perdre). Pour construire vos requêtes vous pouvez utiliser l'éditeur de requête.



1. Sélectionner les communes ayant un nombre d'habitants supérieur à 10 000.
2. Calculer la superficie (Area) totale (Sum) par groupe écologique (FACIES2) (Group By)
3. Calculer la superficie totale des forêts (Sum())
4. Sélectionner les forêts présentes dans des communes habitées par plus de 10 000 habitants
5. Quelle est la commune la plus habitée ? (Utilisez l'agrégation Max() avec une sélection imbriquée ou un tri descendant par population (Order by Population DESC) avec un affichage limité à 1 : LIMIT 1)
6. Quelle est le groupe écologique le plus vaste ? Pour répondre à la question, renommer dans la requête 2 la sum() en ajoutant juste après sum la chaîne "surface" ( sum(...) "surface" ) et dans le order by ajoutez surface (avec DESC et LIMIT éventuellement)