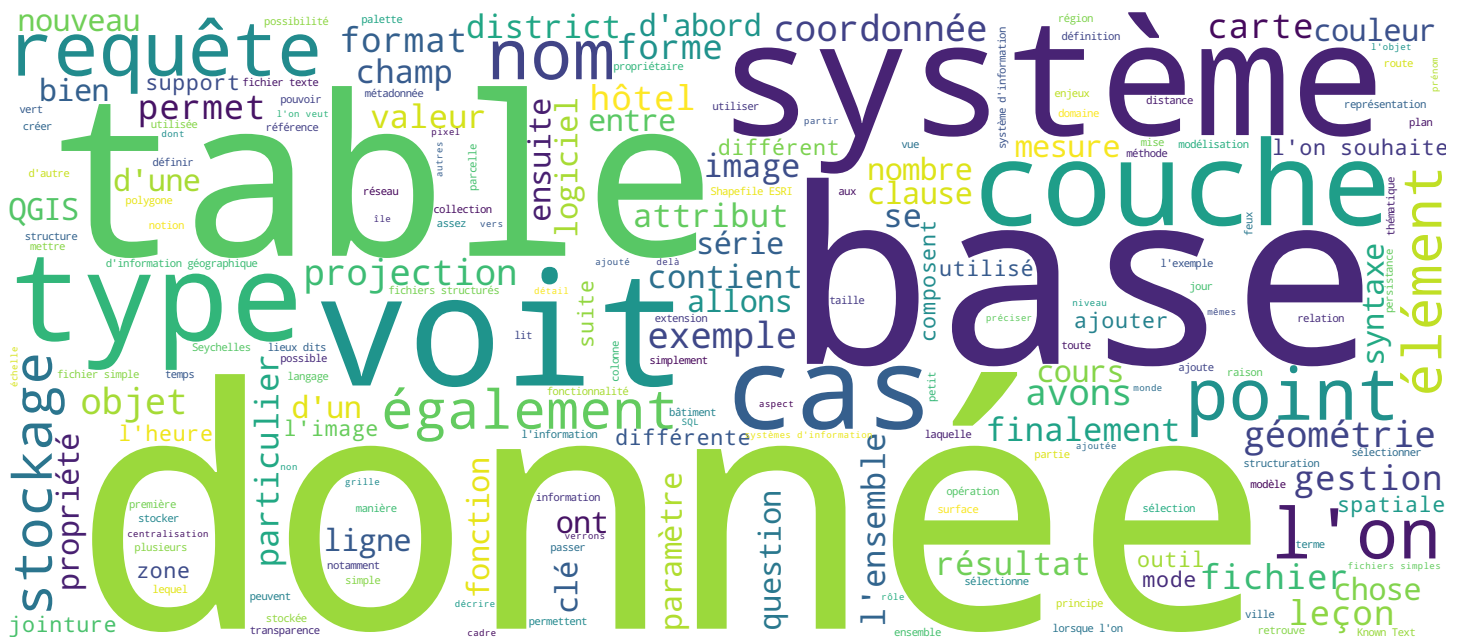


Stockage de l'information géographique

Introduction aux systèmes d'information géographique

Stéphane Joost, Marc Soutter, Fernand Kouamé, Amadou Sall



Search MOOC



Video



Que souhaite-t-on stocker ?



Au niveau de l'objet

- La géométrie: type et coordonnées des sommets

Au niveau de la collection d'objets

- Le système de projection

Introduction aux systèmes d'information géographique

Bienvenue à cette leçon qui portera sur le stockage de données géographiques. En effet, les géodonnées pour pouvoir être utilisées dans les systèmes d'information géographique doivent être stockées sur un support numérique. Et c'est cette question-là que nous allons aborder ce matin. Nous aborderons donc les questions suivantes : que souhaite-t-on stocker ? Quel type d'information ? Quels sont les enjeux du stockage de géodonnées notamment en termes de structuration, d'accessibilité ? Et quels sont les types de supports qui sont les mieux adaptés à répondre à ces différents enjeux ? Comme les choses sont passablement différentes entre le monde vecteur et le monde raster, nous traiterons ces 2 cas successivement et nous allons commencer par le monde vecteur. Nous voyons dans cette image une série d'objets vectoriels, des champs, des routes et une collection d'arbres. L'information la plus fondamentale que l'on souhaite conserver par rapport à ces objets concerne leur géométrie, en particulier le type de géométrie auquel on a affaire et les coordonnées des différents sommets qui la composent. Et qui dit coordonnées, dit également système de projection, donc ces 2 types d'information devraient être conservés.

Notes

Summary



Que souhaite-t-on stocker ?



Au niveau de l'objet

- La géométrie: type et coordonnées des sommets
- Les données attributaires

Au niveau de la collection d'objets

- Le système de projection
- La symbologie
- La topologie (connectivité des objets p. ex.)

Introduction aux systèmes d'information géographique

Au delà de la géométrie, ce qui nous intéresse particulièrement c'est les caractéristiques de ces objets. En particulier, le type de culture dans le cas des champs ou le propriétaire. Ce que l'on résume sous le nom, la dénomination de données attributaires. Les éléments de représentation de ces objets sont également importants en particulier dans le cas du remplissage, la couleur et la transparence, et dans le cas du contour, la couleur, la transparence mais aussi l'épaisseur, tous éléments qui font partie de la symbologie. Et finalement, ce qui nous intéresse également c'est la manière dont les objets sont connectés entre eux, c'est-à-dire la topologie, en particulier dans des systèmes de réseaux. On peut penser par exemple à un réseau de cours d'eau, on aimerait savoir dans quel sens l'eau s'écoule et comment les différentes rivières sont connectées les unes aux autres.

Notes

Summary



1m 42s

Enjeux du stockage de données et types de supports

- Persistance des données
Objectif premier
- Structuration des données
Aspect fondamental

Introduction aux systèmes d'information géographique

Si on se penche maintenant sur les enjeux du stockage de données, et sur le type de support utilisé, on peut relever que l'objectif premier du stockage, c'est la persistance des données. Une donnée est dite persistante si, lorsqu'elle a été ajoutée à un support, elle reste accessible tant qu'elle n'a pas été explicitement supprimée. Au-delà de la persistance, ce qui nous intéresse c'est la structuration des données qui constitue un aspect très fondamental. Ce que l'on peut bien comprendre au travers de l'exemple qui suit.

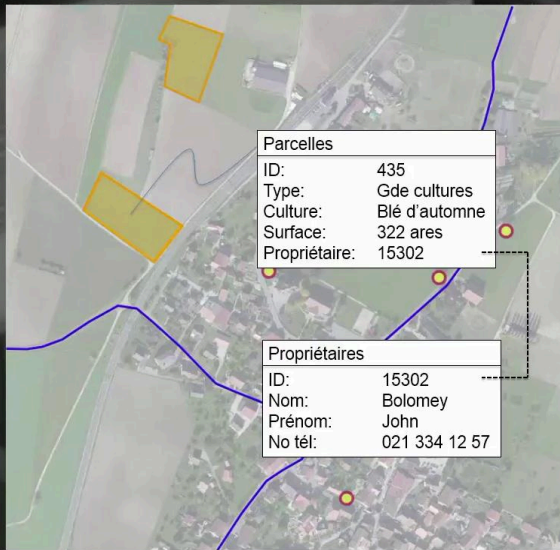
Notes

Summary



2m 44s

Enjeux du stockage de données et types de supports



Un exemple

- Chaque parcelle a un propriétaire.
- Plusieurs parcelles >> même propriétaire.
- Répétition des détails relatifs au propriétaire, >> risques d'erreur lors de mises à jour.
- Séparer données parcelles et propriétaires et créer une relation passant par un identifiant.

➡ Structuration des données.

Introduction aux systèmes d'information géographique

On retrouve sur cette image les parcelles de tout à l'heure avec quelques données supplémentaires concernant leurs propriétaires : le nom, le prénom et le numéro de téléphone. Chaque parcelle a un propriétaire et plusieurs parcelles peuvent-avoir un même propriétaire. Ce qui conduit à répéter des détails relatifs aux propriétaires pour chaque parcelle, avec des risques d'erreurs lors notamment des mises à jour par exemple. Pour éviter ces erreurs, il faut séparer en fait les données concernant les parcelles des données concernant les propriétaires et établir un lien entre ces 2 jeux de données par le biais d'un identifiant. Nous reparlerons plus en détails de cette structuration des données dans la leçon numéro 2 de ce cours qui portera sur précisément la structuration et la modélisation des données.

Notes

Summary



3m 16s

Enjeux du stockage de données et types de supports

- Centralisation de l'accès au données
Sécurité et intégrité des données
- Fonctionnalités de gestion
Ajout, modification, recherche de données à l'aide de logiciels existants

Introduction aux systèmes d'information géographique

Parmi les autres enjeux du stockage des données, il y a la centralisation de l'accès aux données pour des raisons de sécurité. Lorsqu'on a des données confidentielles, on aimerait éviter de les disperser sur un grand nombre de supports dans des lieux différents. Et pour des raisons d'intégrité, plus on multiplie les versions d'une base de données, et plus il y a des risques que ces données deviennent disparates et ne soient plus conciliables. Finalement, les fonctionnalités de gestion sont également un enjeu important puisque l'on aimerait, lorsque l'on veut ajouter, modifier ou rechercher des données dans une base de données, éviter de reprogrammer la chose et utiliser les outils existants dans les logiciels dédiés.

Notes

Summary



4m 09s

Formes de stockage de données et types de supports

- Persistance

Stockage en fichiers simples

- Structuration

Stockage en fichiers semi-structurés, base de données autonomes

- Centralisation des accès

Fichiers ou base de données dans une architecture client-serveur

- Fonctionnalités de gestion

Système de gestion de base de données (SGBD)

Introduction aux systèmes d'information géographique

Si l'on examine maintenant les enjeux du stockage des données en relation avec les différents types de support que l'on peut envisager d'utiliser, on voit que du point de vue de la persistance, le stockage en fichier simple est tout à fait suffisant. Et si l'on souhaite ajouter une possibilité de structurer les données, il faut passer à un format un peu plus élaboré, celui des fichiers structurés. Et nous verrons dans la suite de cette leçon des exemples de fichiers simples et de fichiers structurés. Les éléments base de données pour la structuration seront abordés plus en détail dans la suite du cours. La question de la centralisation des accès nous amène à celle de l'architecture client-serveur, avec une base de données ou des fichiers indépendants, logés, hébergés sur un serveur central accessible par une multitude de clients. La question des fonctionnalités de gestion nous amène au système de gestion de base de données disponible sur le marché.







Notes

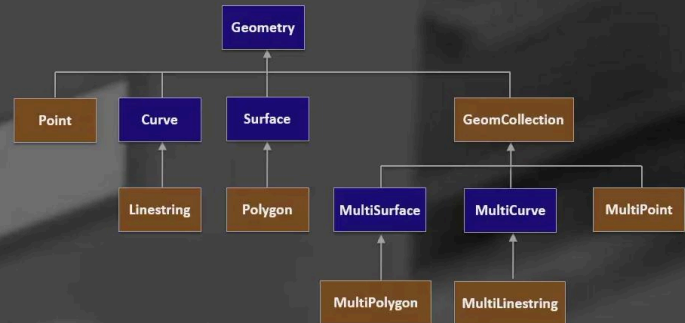
Summary



5m 01s

Stockage en fichiers simples - .wkt Well Known Text

Type	Examples	
Point	POINT(30 10)	
LineString	LINESTRING (30 10, 10 30, 40 40)	
Polygon	POLYGON ((30 10, 10 20, 20 30, 40 30, 30 10)) POLYGON ((35 10, 10 20, 15 40, 40 40, 35 10), (20 30, 35 35, 30 20, 20 30))	
Multipoint	MULTIPOINT ((10 40), (40 30), (20 20), (30 10)) MULTIPOINT (10 40, 40 30, 20 20, 30 10)	
MultiLineString	MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))	
MultiPolygon	MULTIPOLYGON (((30 20, 10 40, 10 40, 45 40, 30 20), (15 5, 40 10, 10 20, 5 10, 15 5))) MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 45 20, 30 5, 10 10, 10 30, 20 35), (30 20, 20 25, 20 15, 30 20)))	



Introduction aux systèmes d'information géographique

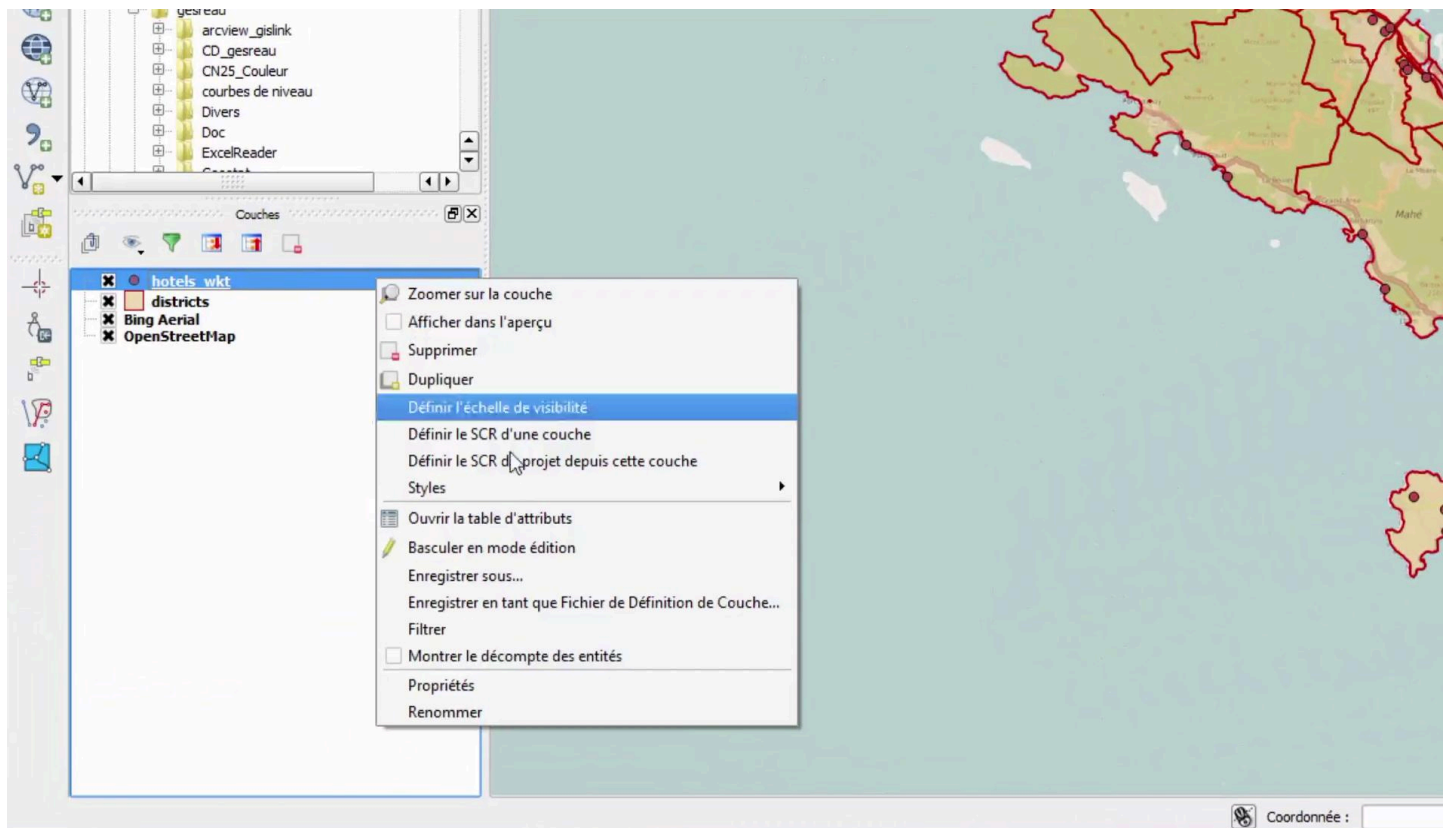
Dans la suite de cette leçon, nous allons maintenant nous concentrer donc sur le stockage en fichier simple et le stockage en fichier structuré en prenant divers exemples de formats qui sont utilisés dans ce cadre. Le premier de ces formats de stockage en fichiers simples connu sous le nom de Well Known Text ou WKT, est un format tout à fait transparent puisqu'il est lisible, c'est un simple fichier texte ASCII, qui contient en fait les géométries décrites par un mot clé : le point, LineString, polygone et puis la succession des coordonnées des différents sommets qui le composent, des paires de coordonnées, des sommets qui le composent. Les 6 formes géométriques de base qui sont listées ici sont généralement reconnues par tous les logiciels système d'information géographique, ce qui rend ce format extrêmement flexible et polyvalent.

Notes

Summary



6m 02s



Dans QGIS le format Well Known Text peut être testé à l'aide d'une extension qui s'appelle QuickWKT. Cette extension propose une fenêtre dans laquelle on peut saisir un String Well Known Text, donc le type de géométrie et les coordonnées des points qui le composent, ici, un point qui se situe quelque part dans l'océan Indien à proximité de l'île de Mahé aux Seychelles. Un point et une ligne qui sont ajoutés en fait à la carte. L'ajout de ces éléments crée de nouvelles couches que l'on peut ensuite supprimer lorsqu'on en a plus besoin. Une autre manière d'utiliser le WKT dans QGIS, c'est de créer un fichier texte dans lequel on va ajouter en fait dans la première ligne les en-têtes des différentes colonnes qui composent les données avec le mot clé WKT qui désigne la colonne contenant les coordonnées des points et puis ensuite les différents attributs de la couche, ici les hôtels des Seychelles. Et on voit maintenant que l'on peut importer cette couche, donc on ajoute une couche de type vectoriel. On sélectionne le format CSV, il faut préciser le système de projection utilisé en l'occurrence UTM 40 Sud et la couche est ajoutée.

Notes

Summary



Stockage en fichiers simples – .shp shapefile ESRI

- Un standard de fait
- En réalité plusieurs fichiers, dont trois sont indispensables
 - .shp** entités géographiques
 - .dbf** données attributaires
 - .shx** index des entités géographiques
- et des fichiers accessoires, comme par exemple
 - .prj** paramètres du système de projection.
 - .sbn, .sbx** index relatifs aux requêtes.

Introduction aux systèmes d'information géographique

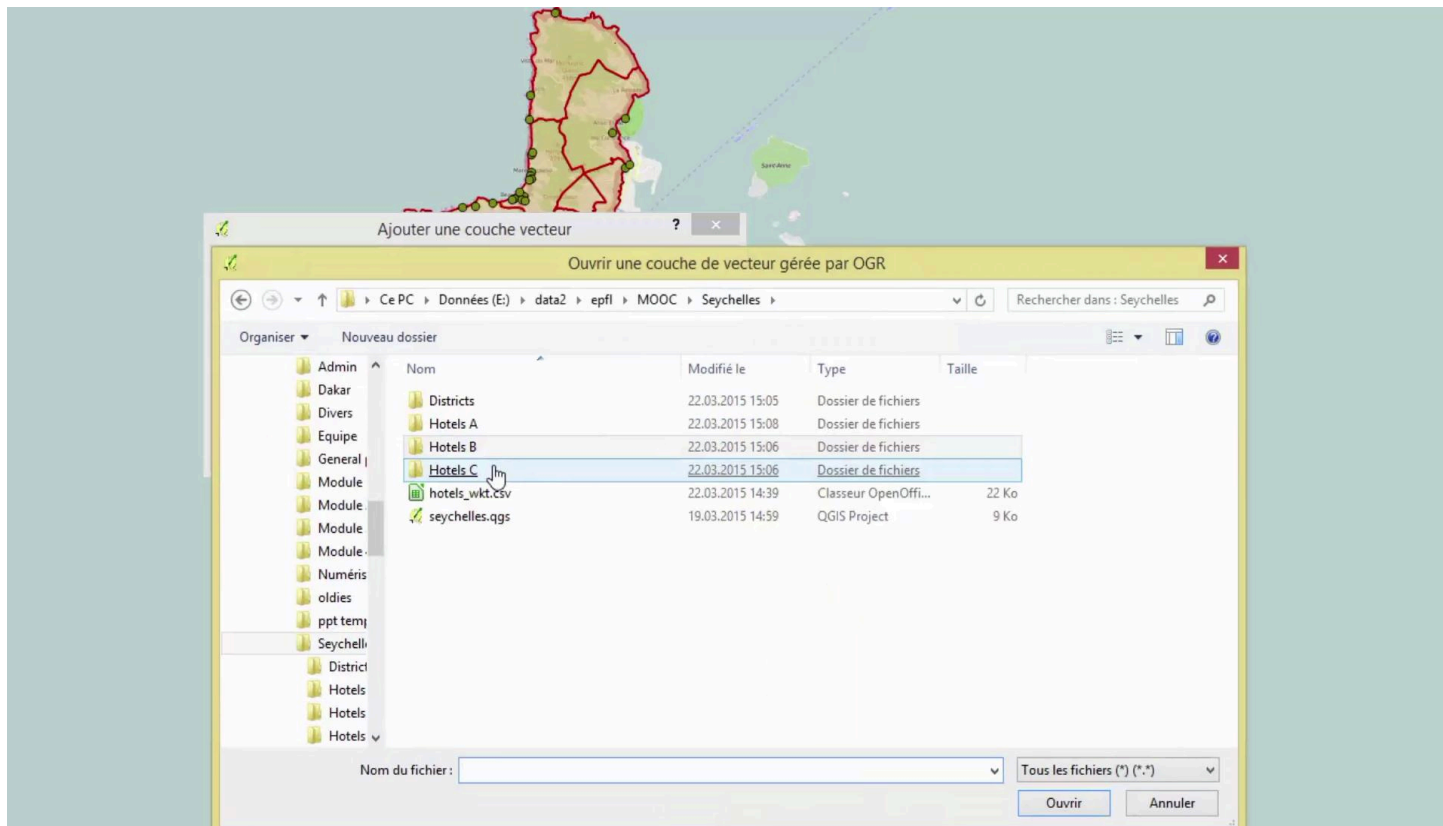
Cette couche peut également être enregistrée au format VKT en précisant un fichier de type CSV, donc séparé par des virgules, et en précisant que la géométrie doit être décrite sous forme VKT. Le deuxième type de format de fichier simple qu'il faut connaître c'est le shapefile ESRI, qui est devenu un standard de fait puisqu'il est utilisé, en tout cas reconnu, par quasiment tous les logiciels de SIG. Ce format shapefile est en réalité constitué de plusieurs fichiers, dont 3 sont indispensables : le.shp qui contient les entités géographiques, le.dbf qui contient les données attributaires et le.shx qui contient un index des entités géographiques. À cela s'ajoute une série de fichiers accessoires, en particulier le.prj qui contient les paramètres du système de projection et d'autres encore.

Notes

Summary



8m 39s



Dans l'exemple qui est montré ici, nous avons préparé plusieurs séries de fichiers Shapefile ESRI. Et dans un premier temps, on montre qu'un fichier dbf peut être ouvert par un logiciel tableur en l'occurrence LibreOffice puisque c'est un fichier en fait qui contient des données sous forme de table attributaire. Dans le deuxième jeu de données on a ajouté le fichier shx, c'est dans le troisième, le fichier prj qui contient donc les paramètres du système de projection. On voit que ensuite dans QGIS, si l'on veut ajouter une couche de données vectorielles de type Shapefile ESRI, on choisit d'ajouter une couche vecteur, et on voit que si on sélectionne le premier cas de figure où il manque le fichier shx, en fait l'importation échoue avec un message d'erreur qui permet de se convaincre que si les 3 fichiers ne sont pas présents ça ne va pas fonctionner. Dans le deuxième exemple ici, on prend le second cas où nous avons les 3 fichiers mais pas le fichier de projection. L'importation se fait mais on doit préciser le système de projection avant de pouvoir ajouter la couche sur la carte. Et finalement dans le troisième exemple, lorsque l'on a les 3 fichiers de base plus le fichier de projection, qui est l'exemple C, on voit que l'importation se fait directement dans le logiciel.

Notes

Summary



9m 36s

Stockage en fichiers simples – .dxf autocad

- Standard dans le monde des architectes et urbanistes
- Format de DAO
- Ni attributs, ni système de projection, ni symbologie, ni....
- Uniquement des géométries simples (point, lignes, polygones) et des écritures (tout est stocké en vrac)

Introduction aux systèmes d'information géographique

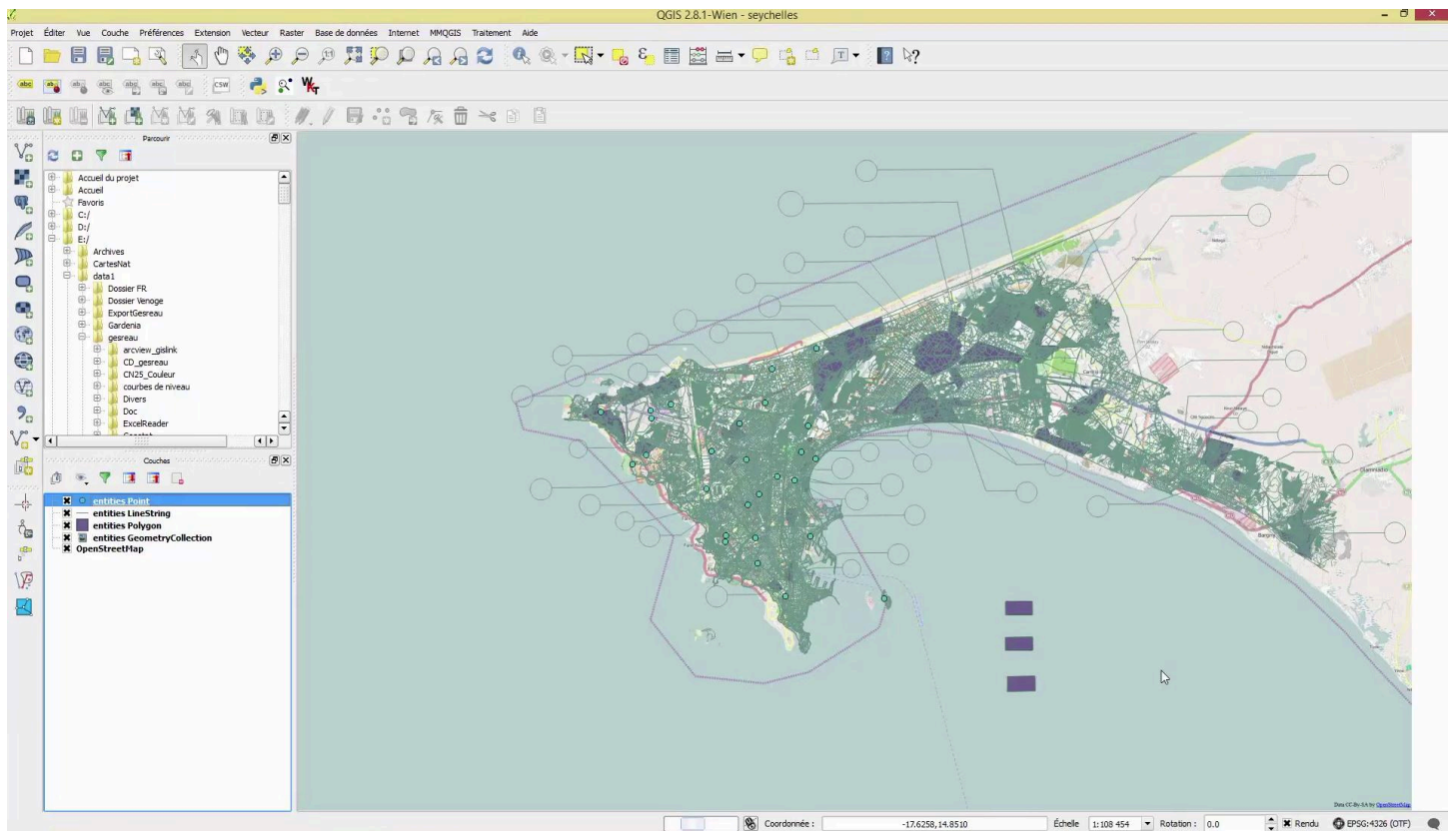
Le troisième type de format de fichier standard simple que l'on souhaitait aborder aujourd'hui, c'est le dxf autocad qui est un format également très répandu, puisque ça constitue un standard dans le monde des architectes et des urbanistes. Il est important de savoir qu'il s'agit non pas d'un format de fichier SIG, mais d'un format de DAO, donc Dessin Assisté par Ordinateur. Et en fait ce format ne contient ni attribut, ni système de projection, ni symbologie. En fait il se résume uniquement à des géométries simples, à des points, des lignes, des polygones et à des écritures, le tout étant stocké en vrac dans un seul fichier.

Notes

Summary



11m 11s



Dans cet extrait vidéo, on montre tout d'abord qu'un fichier dxf est un fichier D5 qui peut être ouvert par un logiciel de dessin, en l'occurrence LibreOffice de nouveau, et on voit qu'ici pour la région de Dakar, si on zoom sur la région de l'aéroport, on a l'ensemble des éléments du plan directeur de la région. L'importation d'un fichier dxf dans QGIS passe par la définition du système de projection et puis par la sélection en fait du type d'entité présent dans le fichier et que l'on souhaite importer si l'on choisit d'importer tout : points, lignes, polygones, écritures et puis ces éléments s'affichent en vrac.

Notes

Summary



11m 56s

Stockage en fichiers structurés



Ferme
ID
Exploitant
Tel.



Champ
ID
culture
surface
plantation

Bâtiment
nom
description
adresse
surface

Fichiers simples

2 fichiers géométrie + attributs

- Champ (.wkt, .shp)
- Bâtiment (.wkt, .shp)

1 fichier d'attributs

- Ferme (.csv, .dbf)

Fichiers structurés

1 structure hiérarchisée

```
<Fermes>
  <Ferme>
    <Bâtiments>
      ...
    </Bâtiments>
    <Champs>
      ...
    </Champs>
  </Ferme>
  ...
</Fermes>
```

Introduction aux systèmes d'information géographique

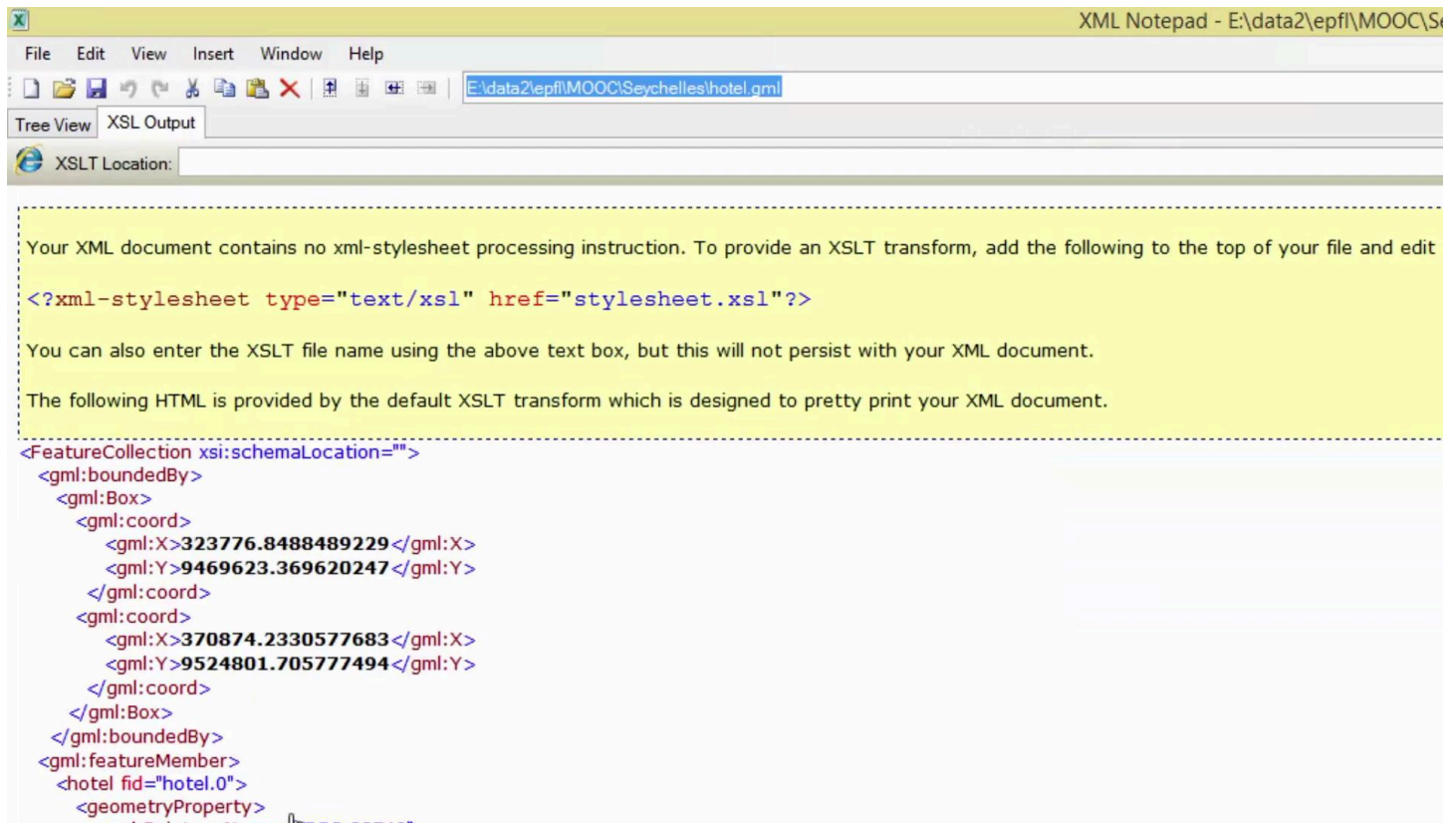
Nous abordons maintenant la question du stockage dans les fichiers structurés. Nous retrouvons l'image de tout à l'heure avec une ferme composée de plusieurs parcelles, de plusieurs champs et d'une série de bâtiments, chacun avec des caractéristiques différentes. Dans une approche basée sur des fichiers simples, on utiliserait 2 fichiers de type Well Known Text ou shapefile ESRI pour stocker la géométrie et les attributs des champs et des bâtiments et un fichier d'attribut pour caractériser les différentes fermes, qui serait un fichier de type.csv ou.dbf. Une alternative à cette approche consiste à utiliser des fichiers structurés, de type XML, donc des fichiers par balise où dans l'exemple que l'on voit ici, on a une collection de fermes en jaune qui contient des objets fermes contenant eux-même des collections de bâtiments et des collections de champs. Et de manière hiérarchisée, on aboutit en fait à intégrer l'ensemble de l'information dans une seule structure de données. Nous allons voir maintenant 2 exemples de fichiers de type structuré.

Notes

Summary



12m 49s



Nous retrouvons dans QGIS la couche des hôtels de l'île de Mahé que nous exportons, on choisit un format géographique Markup Language, format GML qui est un format de fichier structuré, on corrige 2 ou 3 paramètres d'exportation et puis dans la foulée, on exporte la même couche de données dans un autre format structuré qui est le GeoJSON également sous le nom d'hôtel avec une extension cette fois différente. Si l'on regarde maintenant à quoi ressemblent ces fichiers en commençant par le GeoJSON... à l'aide d'un éditeur de fichier JSON de type, on voit qu'on a une structure hiérarchique avec tout d'abord les propriétés du système de projection puis les différents objets donc la géométrie et les coordonnées de l'hôtel, puis les différents attributs en listes. Cette représentation hiérarchique peut aussi être visualisée sous format texte avec pour chaque élément en fait les informations stockées sur une seule ligne. Le fichier GML a une structure similaire, c'est donc un dérivé du XML. Et on peut également le consulter sous forme de structure hiérarchique avec ici d'abord le bounding box, avec les coordonnées de l'ensemble de la couche. Et puis pour chaque élément, les différentes caractéristiques, géométries, attributs, etc.

Notes

Summary

14m 08s



XML Notepad - E:\data2\epfi\MOOC\Sc

File Edit View Insert Window Help

E:\data2\epfi\MOOC\Seychelles\hotel.gml

Tree View XSL Output

XSLT Location:

<gml:X>357251.15758085577</gml:X>
<gml:Y>9469623.369620247</gml:Y>
</gml:coord>
<gml:coord>
<gml:X>370874.2330577683</gml:X>
<gml:Y>9524801.705777494</gml:Y>
</gml:coord>
</gml:Box>
</gml:boundedBy>
<gml:featureMember>
<hotel fid="hotel.0">
<geometryProperty>
<gml:Point srsName="EPSG:32740">
<gml:coordinates>357251.15758085577,9521603.379920898</gml:coordinates>
</gml:Point>
</geometryProperty>
<ID>1</ID>
<NAME>Beach Bungalows</NAME>
<ROOMS>4</ROOMS>
<BEDS>8</BEDS>
<STATUS>Small Hotel</STATUS>
<MANAGER>Mr. A. Confait</MANAGER>
<ADDRESS>Grand' Anse</ADDRESS>
<PHONE>233445</PHONE>
<FAX>233098</FAX>
<E_MAIL>martin@seychelles.net</E_MAIL>
<TELEPHONE>0</TELEPHONE>

Et comme tout à l'heure, on peut également switcher vers une version texte de ce fichier qui reproduit donc le langage par balise où on retrouve notamment la géométrie de l'objet.

Notes

Summary

15m 39s



Que souhaite-t-on stocker ?

No Colonne →

No Ligne ↓



Au niveau de la cellule

- Les coordonnées de la cellule
- La valeur de la variable

Au niveau de la grille

- Le système de projection

Table de correspondance

1	Routes
2	Champs
3	Bâtiments

Cellule ou pixel: Ligne, Colonne, Valeur

Introduction aux systèmes d'information géographique

Il est temps de passer maintenant au mode raster. Les coordonnées XY de l'origine de la grille et la taille des pixels permettent de stocker les coordonnées de la cellule auxquelles il faut ajouter les paramètres du système de projection bien évidemment. Les valeurs de la grille elle-même au travers d'une table de correspondance fournissent les valeurs de la variable finale que l'on souhaite enregistrer.

Notes

Summary

16m 00s



Enjeux et formes de stockage de données



- **Persistence**
Stockage en fichiers simples, images ou tables
- **Structuration**
Images ou tables X,Y,Z simples
- **Centralisation des accès**
Fichiers simples ou tables simples d'une base de données, dans une architecture client-serveur
- **Fonctionnalités de gestion**
Système de gestion de base de données (SGBD)

Introduction aux systèmes d'information géographique

Dans le cas du monde raster, les enjeux de stockage des données sont les mêmes que pour le monde vectoriel, mais les conséquences et le type de support sont assez différents. Du point de vue de la persistance, en fait le stockage en fichier simple sous forme d'image ou de grille fait tout à fait l'affaire. La question de la structuration ne se pose pas vraiment dans la mesure où on a des fichiers ou des structures XYZ qui sont simples. L'aspect de la centralisation des accès joue un rôle. C'est vrai qu'il est intéressant d'avoir ces fichiers rasters stockés sur un serveur dans une architecture client-serveur. Et puis finalement du point de vue fonctionnalité de gestion, on a des systèmes de gestion de base de données qui offrent des fonctionnalités spécifiques à la gestion des rasters et c'est là quelque chose qui est très intéressant.

Notes

Summary



16m 36s

Formats grille



	5	15	25	35
55	NA	NA	14	2
45	NA	21	65	50
35	4	11	31	65
25	38	49	29	13
15	12	36	45	23
5	10	22	8	NA

```
ncols 4
nrows 6
xllcorner 0.0
yllcorner 0.0
cellsize 10.0
nodata_value -9999
-9999 -9999 14 2
-9999 21 65 50
4 11 31 65
38 49 29 13
12 36 45 23
10 22 8 -9999
```

```
DSAA
4 6
5.0 35.0
5.0 55.0
2 65
10 22 8 -9999
12 36 45 23
38 49 29 13
4 11 31 65
-9999 21 65 50
-9999 -9999 14 2
```

```
5 55 -9999
15 55 -9999
25 55 14
35 55 2
5 45 -9999
15 45 21
25 45 65
35 45 50
5 35 4
15 35 11
25 35 31
35 35 65
5 25 38
15 25 49
25 25 29
35 25 13
5 15 12
15 15 36
25 15 45
35 15 23
5 5 10
15 5 22
25 5 8
35 5 -9999
```

Formats matriciels – En-tête + contenu

- Grille Esri ASCII (.asc)
- Grilles Surfer ASCII et binaire (.grd)
- Grille Esri binaire (.flt, .hdr)

Formats «liste» – XYZ

- Fichiers ASCII (.txt, .csv, .dat)

Introduction aux systèmes d'information géographique

Les formats grilles utilisés pour le stockage de rasters se composent de formats matriciels, qui en fait sont constitués d'un en-tête suivi d'un contenu. On a ici des exemples d'une grille ASCII ESRI, avec ensuite une grille Surfer également sous format ASCII, donc en fichier texte. On voit que les mêmes données sont présentées sous forme matricielle, après un en-tête de 5 ou 6 paramètres qui caractérisent en fait la taille de la grille, les coordonnées du point origine, etc. La format liste XYZ, donc 3 colonnes simplement pour décrire l'ensemble de ces données dans un format.txt,.csv,.dat est aussi un grand classique.

Notes

Summary



Formats image

Formats matriciels - En-tête et contenu

- Formats image .tif, .jpg, .png, .gif

L'information colorimétrique peut être stockée sous la forme:

- des 3 (ou 4) composantes de base de chaque pixel (RVB ou RVBA),
- d'une palette des couleurs présentes dans l'image et, pour chaque pixel, du numéro d'index de sa couleur dans la palette.

Introduction aux systèmes d'information géographique

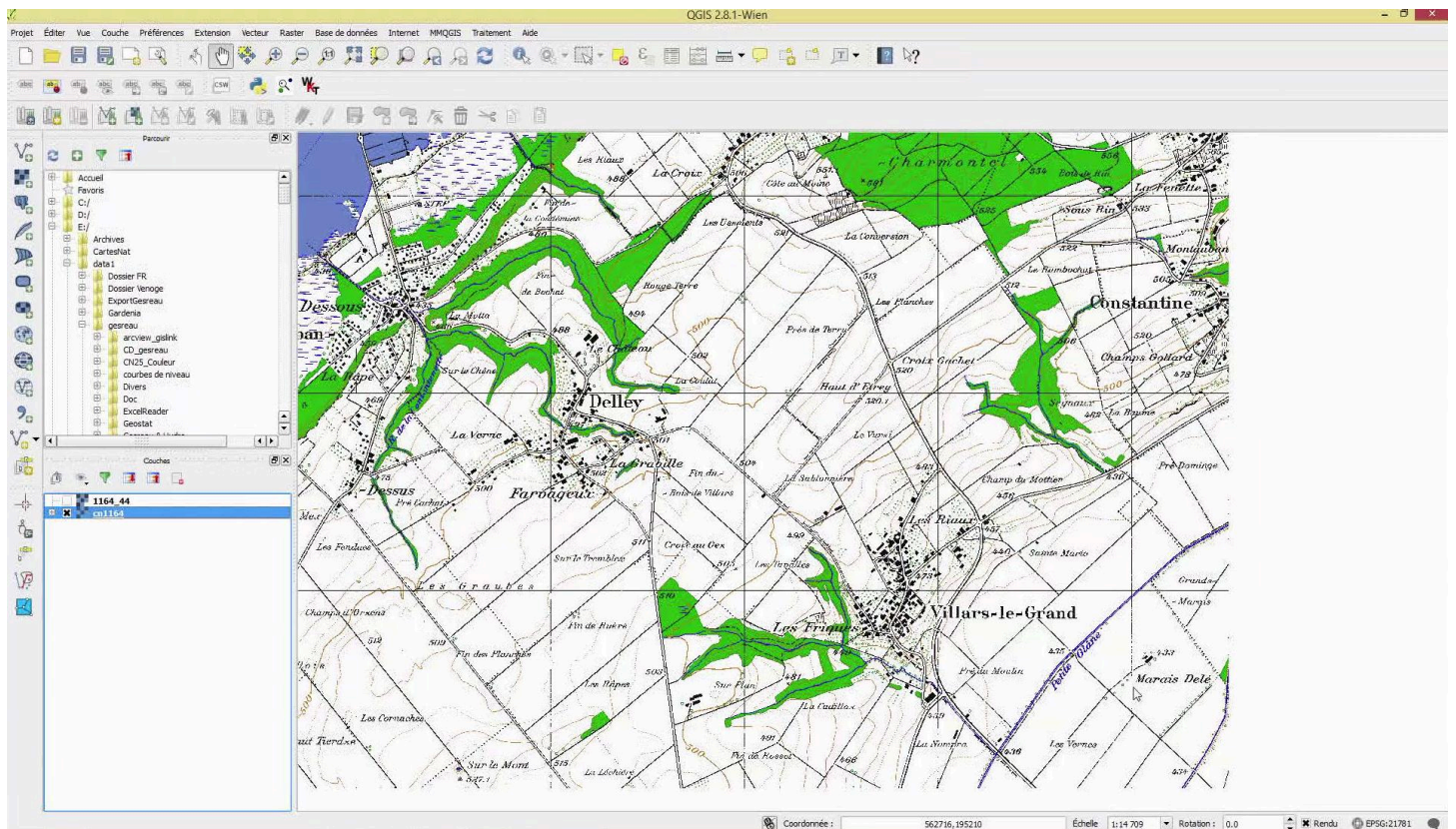
Les formats images sont également des formats matriciels, constitués d'un en-tête contenant les paramètres descriptifs de l'image suivi d'un contenu sous forme de tableau. Les principaux formats images utilisés sont le tif, le jpg, le png et le gif. L'information colorimétrique peut être stockée sous 2 formes : soit les différentes composantes de base de chaque pixel rouge, vert, bleu ou éventuellement la transparence sont stockées, soit on stocke une palette de l'ensemble des couleurs présente dans l'image et pour chaque pixel, on stocke le numéro d'index de sa couleur dans la palette.

Notes

Summary



18m 31s



Nous avons ici un exemple de deux couches rasters, une image et un fond de carte qui illustrent ces 2 modes de stockage de l'information colorimétrique. Tout d'abord pour l'image si l'on regarde ses propriétés, on voit que nous avons une série de bandes correspondant au rouge, au vert et au bleu qui composent la colorimétrie de l'image et si l'on désactive l'une ou l'autre de ces couleurs, en fait seul restent les 2 autres. Et cela permet de voir qu'en fait on a dans ce type de format de stockage, en fait 3 images superposées qui donnent l'image de couleur finale. Dans le cas du fond de plan, on a l'autre alternative, une palette de couleurs, dont seules les premières sont en fait utilisées et on voit qu'on peut changer l'une de ces couleurs, ici le vert des forêts pour un vert un peu fluo vers un vert plus sombre que l'on applique ou le bleu du lac également dans une teinte plus sombre. Ce qui permet de manipuler en fait de manière globale la palette de couleurs de cette carte. Finalement, l'information de positionnement de l'image peut se trouver soit dans l'en-tête d'un fichier tif et on parle alors de format GeoTiff, soit dans un world file qui accompagne le fichier image avec les extensions tfw, jpgw, etc.

Notes

Summary

19m 10s

