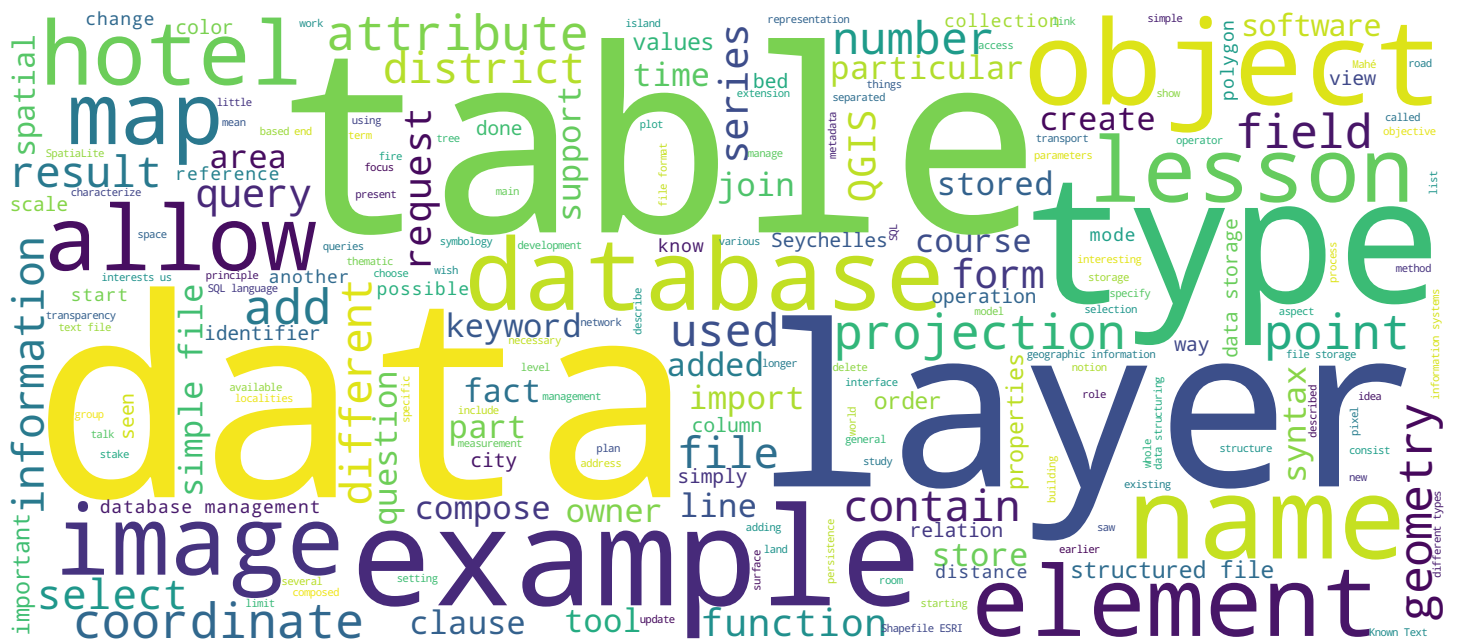


Storage of geographic information

Stéphane Joost, Marc Soutter, Fernand Kouamé, Amadou Sall

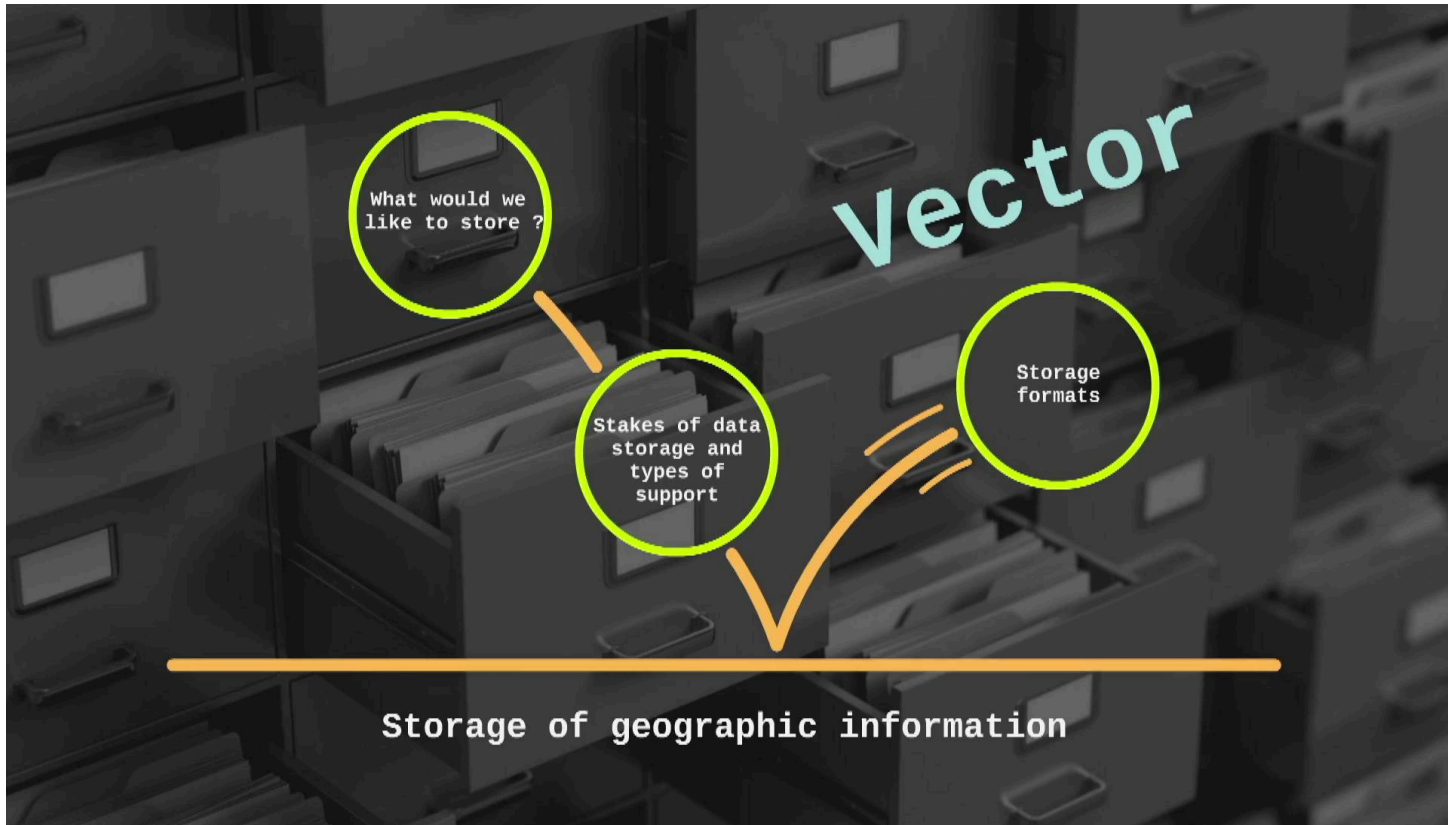


Search MOOC



Video





Welcome to this lesson which will focus on geographic data storage. Indeed, to be able to use geodata in geographic information systems, they must be stored on a digital support. And this is the question we will be discussing this morning. We will therefore address the following questions: what do we want to store? What kind of information? What are the stakes involved in the storage of geodata particularly in terms of structuring, accessibility? And what are the types of supports that are best suited to meet these challenges? As things are quite different between the vector world and the raster world we will treat these two cases successively and we will start with the vector world.

Notes

Summary



0m 23s

Stakes of data storage and types of support

- Data persistency
Primary goal
- Data structuration
Fundamental aspect

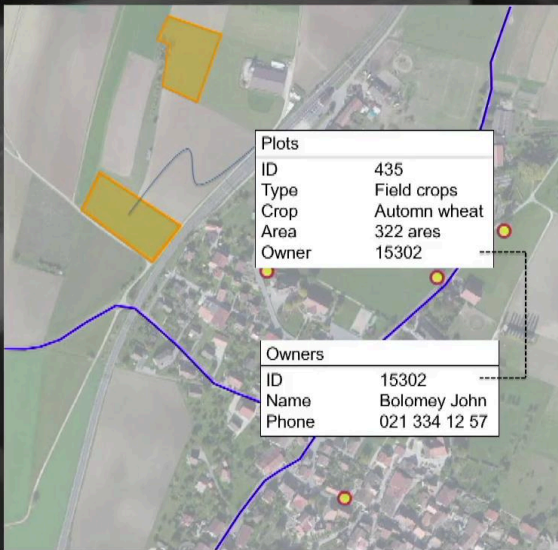
We see in this image a series of vectorial objects, fields, roads and a collection of trees. The most fundamental information that we wish to keep with regards to these objects concerns their geometry, in particular the type of geometry we are facing and the coordinates of the different summits that compose it. And who say coordinates also says projection system, so these two types of information should be kept. Beyond geometry, what particularly interests us is the characteristics of these objects. Particularly, the type of crop in the case of fields or the owner. What is summarized under the name, the denomination of attribute data. The elements of representation of these objects are also important in particular in the case of the filling, the color and the transparency, and in the case of the contour, the color, the transparency but also the thickness, all the elements which are part of the symbology. And finally, what interests us also is the way the objects are connected to each other. that is to say the topology, in particular in network systems. For example we can think of a stream network, we would like to know in what direction the water flows and how the different rivers are connected to each other.

Notes

Summary



Stakes of data storage and types of support



An example

- Each plot has an owner
- Several plots >> same owner
- Repetition of the details related to the owner >> risk of errors during updates
- Separate plots' data and owners' data and create a relation through an identifier

➔ Data structuration

An Introduction to Geographic Information Systems

Looking now at the stakes of data storage, and at the type of media used, we notice that the primary objective of the storage is the persistence of the data. A data is said to be persistent if, when added to a support, it remains accessible, as long as it has not been explicitly deleted. Beyond the persistence, what interests us is the data structuring which constitutes a very fundamental aspect. We will understand it through the following example. On this image we can see the plots seen earlier with some additional data about their owners: the name, first name and phone number. Each plot has an owner and several plots can have the same owner. Which leads to repeat details relating to owners for each plot, with error risks especially during updates for example. In order to avoid these errors, the data regarding the plots must be separated from the data regarding the owners and a link between these two sets of data must be established by means of an identifier. We will talk more about this data structuring in lesson no. 2 of this course which will focus precisely on the data structuring and data modeling.

Notes

Summary



2m 44s

Stakes of data storage and types of support

- Centralized access to data
Data security and integrity
- Management functionalities
Data input, modification, or
query using existing software

Amongst the other stakes of data storage, there is the centralization of the data access for security reasons. When we have confidential data, we would like to avoid dispersing them on a large number of supports in different places. And for integrity reasons, the more we multiply the versions of a database the more risks there are that these data become disparate and can no longer be compatible. Finally, the management functionalities are also an important stake since we would like to avoid reprogramming the whole thing and use existing tools in dedicated softwares when we want to add, modify and search data in a database.







Notes

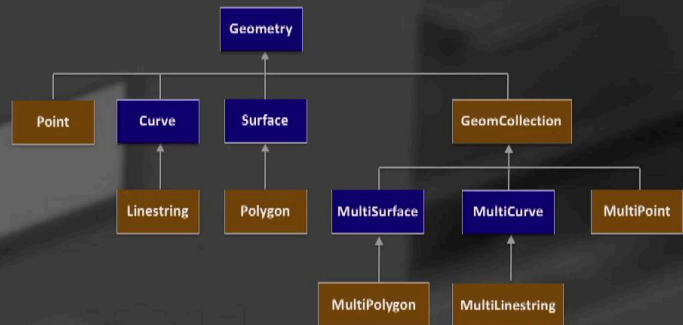
Summary



4m 08s

Simple file storage – Well Known Text, .wkt

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)))	



An Introduction to Geographic Information Systems

If we now examine the stakes of data storage in relation to different types of support that we can envisage to use, we see that from the perspective of the persistence, the storage in simple file is quite sufficient. And if we want to add a possibility to structure the data, we must move to a more elaborate format than that of the structured files. And in the rest of this lesson we will see examples of simple files and structured files. The database elements for structuring will be discussed in more details later in the course. The question of the centralization of access leads us to that of the client-server architecture, with a database or independent files, housed, hosted on a central server accessible by a multitude of clients. The question of management functionalities brings us to the database management system available on the market. In the rest of this lesson, we will now concentrate on the simple file storage and the structured file storage by taking various examples of formats that are used in this framework.







Notes

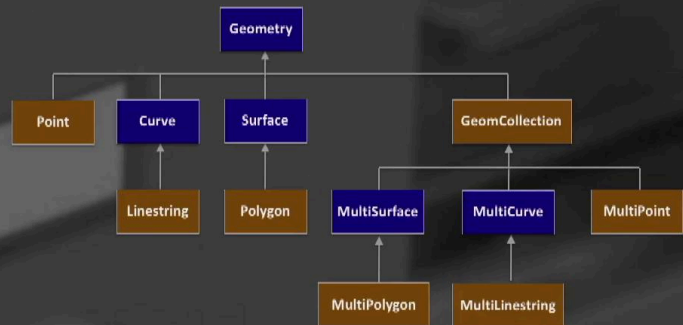
Summary



5m 01s

Simple file storage – Well Known Text, .wkt

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)))	



An Introduction to Geographic Information Systems

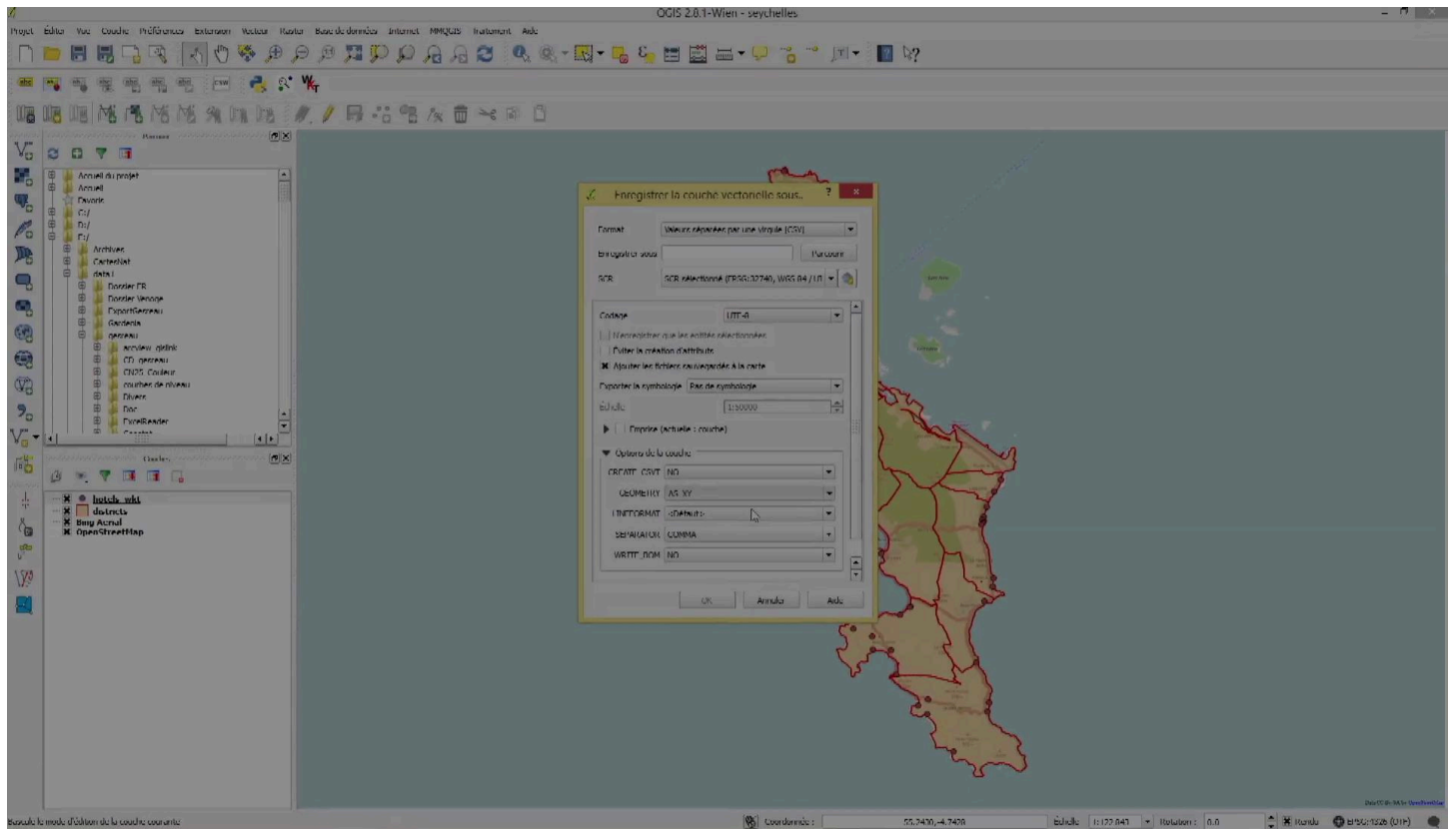
The first of these simple file storage formats known as the Well Known Text or WKT, is a transparent format since it is readable, it is a simple ASCII text file, which actually contains the geometries described by a keyword: the point, LineString, polygon and then the succession of coordinates of the different summits that compose it, the pairs of coordinates of the summits that compose it. The 6 basic geometric forms that are listed here are generally recognized by all the geographical information system softwares. This makes this format extremely flexible and versatile.

Notes

Summary



6m 14s



In QGIS, the Well Known Text format can be tested using an extension called QuickWKT. This extension proposes a window in which we can enter a String Well Known Text, so the geometry type and the coordinates of the points that compose it, here, a point located somewhere in the Indian Ocean near the island of Mahé in the Seychelles. A point and a line that are added in fact to the map. The addition of these elements creates new layers which we can then delete when we no longer need them. Another way to use the WKT in QGIS, is to create a text file in which we will add in the first line the headings of the different columns that compose the data with the keyword WKT that indicates the column containing the coordinates of the points and then the different attributes of the layer, here the Seychelles hotels. And we now see that we can import this layer, so we add a layer of vectorial type. The CSV format is selected, it is necessary to specify the projection system used in this case UTM 40 South and the layer is added. This layer can also be saved in VKT format by specifying a file of CSV type, so separated by commas. and by specifying that the geometry has to be described in VKT form.

Notes

Summary



Simple file storage – ESRI shapefile, .shp

- An actual standard
- In fact several files, of which three are compulsory
 - .shp geographic feature
 - .dbf attribute data
 - .shx positional index of the feature geometry
- Along with subsidiary files, such as
 - .prj coordinate system and projection information
 - .sbn,.sbx spatial index

An Introduction to Geographic Information Systems

The second type of simple file format that we should know is the Shapefile ESRI which has become a de facto standard since it is used, at least recognized, by almost all GIS softwares. This shapefile format is actually made up of several files, 3 of which are essential: the.shp that contains the geographical entities, the.dbf which contains the attribute data and the.shx which contains an index of geographic entities A series of accessory files is added especially the.prj which contains the settings of the projection system and others.

Notes

Summary



8m 57s

Simple file storage – Autocad, .dxf

- Widespread among architects and urbanists
- CAD Format
- No attributes, no projection system, no symbology, neither....
- Only simple geometries (points, lines, polygons) along with writings (bulk storage)

An Introduction to Geographic Information Systems

In the example shown here, we have prepared several series of shapefile ESRI files. First, we show that a dbf file can be opened by a spreadsheet program in this case OpenOffice since it is a file that contains data in the form of an attribute table. In the second dataset the shx file was added, it is in the third one, the prj file which contains the settings of the projection system. We see next that in QGIS, if we want to add a layer of vectorial data of Shapefile ESRI type, we choose to add a vector layer, and we see that if we select the first case where the shx file is missing, the import fails with an error message which convinces us that if the 3 files are not present it will not work. In the second example here, we take the second case where we have the 3 files but not the projection file. The import is done but we have to specify the projection system before adding the layer on the map. And finally in the third example, when we have the 3 basic files plus the projection file which is example C, we see that the import is done directly in the software. The third type of simple standard file format we wanted to discuss today, is the dxf autocad which is also a very popular format, since it is a standard in the world of architects and urban planners.

Notes

Summary



9m 37s

Simple file storage – Autocad, .dxf

- Widespread among architects and urbanists
- CAD Format
- No attributes, no projection system, no symbology, neither....
- Only simple geometries (points, lines, polygons) along with writings (bulk storage)

An Introduction to Geographic Information Systems

It is important to know that this is not a GIS file format, but a CAD format, so Computer Aided Design. And this format does not contain any attribute, no projection system, no symbology. In fact it comes down only to simple geometries, points, lines, polygons and writings, the whole thing being stored in bulk in a single file.

Notes

Summary



11m 28s

Semi-structured file storage



Farm
ID
Owner
Tel.

Field	Building
ID	name
crop	description
area	address
plantation	area

Simple files

2 geometry+attribute files

- Field (.wkt, .shp)
- Building (.wkt, .shp)

1 attribute file

- Farm (.csv, .dbf)

Semi-structured files

1 hierarchized structure

```
<Farms>
  <Farm>
    <Buildings>
      ...
    </Buildings>
    <Fields>
      ...
    </Fields>
  </Farm>
  ...
</Farms>
```

An Introduction to Geographic Information Systems

In this video extract, we first show that a dxf file is a D5 file which can be opened by drawing software in this case OpenOffice again and we see that here in the Dakar region, if we zoom in on the airport area, we have all the elements of the master plan of the region. Importing a dxf file into QGIS goes through the definition of projection system and then the selection of the type of entity present in the file and that we wish to import if we choose to import all: points, lines, polygons, writings and then these items appear in bulk. We now turn to the question of storage in structured files. We can see the image from earlier with a farm composed of several plots, several fields and a series of buildings, each of them with different characteristics. In an approach based on simple files, we would use 2 files of Well Known Text or shapefile ESRI type to store the geometry and the field and building attributes and an attribute file to characterize the different farms, which would be a.csv or.dbf type of file. An alternative to this approach consists in using structured files, such as XML, so tag files where in the example shown here, we have a collection of farms in yellow which contains firm objects containing themselves building collections and field collections.

Notes

Summary



11m 57s

Semi-structured file storage



Simple files

2 geometry+attribute files

Major roads
Building (with .shp)

1 attribute file

Farm (.csv, .dbf)

Semi-structured files

1 hierarchized structure

<Farms>

<Farm>

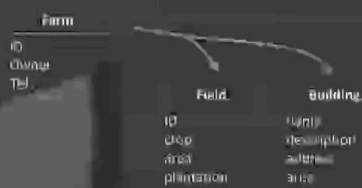
<Buildings>

</Buildings>

</Farm>

</Farms>

</Farms>



And in a hierarchical way, we actually manage to integrate all the information in a single data structure. We will now see two examples of files of a structured type.

Notes

Summary



13m 52s

Semi-structured file storage



Farm
ID
Owner
Tel.

Field	Building
ID	name
crop	description
area	address
plantation	area

Simple files

2 geometry+attribute files

- Field (.wkt, .shp)
- Building (.wkt, .shp)

1 attribute file

- Farm (.csv, .dbf)

Semi-structured files

1 hierarchized structure

```
<Farms>
  <Farm>
    <Buildings>
      ...
    </Buildings>
    <Fields>
      ...
    </Fields>
  </Farm>
  ...
</Farms>
```

An Introduction to Geographic Information Systems

We find in QGIS the layer of hotels in the island of Mahé that we export, we select a Markup Language geographic format, GML format which is a structured file format, we correct 2 or 3 export parameters and then in the process, we export the same data layer in another structured format that is GeoJSON also under the hotel name with a different extension this time. If we now look at what these files look like starting with the GeoJSON... using a file editor of GeoJSON type, we see that we have a hierarchical structure with first the properties of the projection system then the various objects so geometry and the coordinates of the hotel, then the different attributes in lists. This hierarchical representation can also be visualized in text format with for each element the information stored on a single line. The GML file has a similar structure, so this is an XML by-product. And we can also consult it in a hierarchical structure form with here first the bounding box, with the coordinates of the entire layer. And then for each element, the different characteristics, geometries, attributes, etc. And as earlier, we can also switch to a text version of this file which reproduces the language by tag where we find the geometry of the object.

Notes

Summary



14m 08s

What should be stored ?



At the cell level

- Cell coordinate (row/column)
- Value of the thematic variable

At the grid level

- Projection system

Correlation table

1 Roads
2 Fields
3 Buildings

Cell or pixel: Row, Column, Value

An Introduction to Geographic Information Systems

It is now time to switch to the raster mode. The XY coordinates of the grid origin and the pixels size allow to store the cell coordinates to which the parameters of the projection system must be added of course. The values of the grid itself through a correspondence table provide the values of the final variable that we wish to save.

Notes

Summary



16m 00s

Stakes of data storage and types of support

- Persistence
Storage in simple files, images or tables
- Structuration
Images or tables, or X,Y,Z simple files
- Centralized access
Simple files or simple tables in a database, in a client server architecture
- Management functionalities
Database Management System (DBMS)

An Introduction to Geographic Information Systems

In the case of the raster world, the stakes of data storage are the same as for the vectorial world but the consequences and the type of support are quite different. From the persistence point of view, the storage in simple file in the form of an image or a grid perfectly does the trick. The question of structuring does not really arise since we have files or simple XYZ structures. The aspect of centralized access plays a role. It is true that it is interesting to have these raster files stored on a server in a client-server architecture. And finally from a management functionality point of view, we have database management systems that offer functionalities that are specific to the management of raster and this is something that is very interesting.

Notes

Summary



16m 36s

Grid formats



	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
```

Matrix formats – Header + content

- Esri ASCII grid (.asc)
- Surfer ASCII and binary grid (.grd)
- Esri binary grid (.flt, .hdr)

«list» formats – XYZ

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

An Introduction to Geographic Information Systems

The grid formats used for raster storage consist of matrix formats, which are composed of a header followed by a content. Here we have examples of an ASCIIESRI grid followed by a Surfer grid also in ASCII format, so in text file. We see that the same data are presented in a matrix form, after a header 5 or 6 settings which actually characterize the size of the grid, the coordinates of the origin point, etc. The XYZ list format, so 3 columns simply to describe all of these data in a.txt,.csv,.dat format is a classic.

Notes

Summary



Image formats

Matrix formats - Header and content

- Image formats .tif, .jpg, .png, .gif

Colorimetric information may be stored as

- The 3 (or 4) base components of each pixel (RGB or RGBA),
- A color palette of the image's colors, and for each pixel the index number of its color in the color palette

An Introduction to Geographic Information Systems

The image formats are also matrix formats, consisting of a header containing the descriptive parameters of the image followed by a content in the form of a table. The main image formats used are the tif, jpg, png and gif. The color information can be stored in two forms: either the different basic components of each red, green, blue pixel or possibly the transparency are stored, or we store all the colors present in the image and for each pixel, we store the index number of its color in the palette.

Notes

Summary



18m 31s

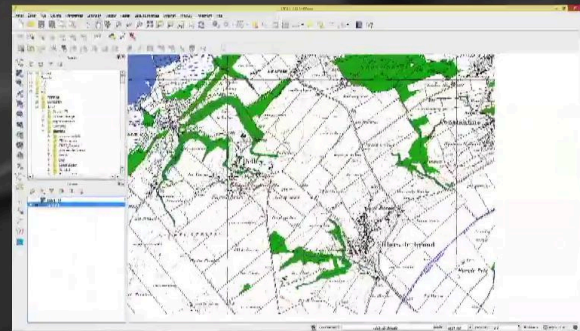
Image formats

Matrix formats - Header and content

- Image formats .tif, .jpg, .png, .gif

Colorimetric information may be stored as

- The 3 (or 4) base components of each pixel (RGB or RGBA),
- A color palette of the image's colors, and for each pixel the index number of its color in the color palette



An Introduction to Geographic Information Systems

Here we have an example of two raster layers, an image and a background map (?) that illustrate these two storage modes of the color information. First for the image if we look at its properties, we see that we have a series of bands corresponding to red, green and blue which compose the colorimetry of the image and if you disable one of these colors, only the other 2 remain. And this allows to see that we have in this type of storage format, 3 superimposed images that give the final color image. In the case of the map (?), we have the other alternative, a color palette of which only the first ones are actually used. and we see that we can change one of these colors, here the green of the forest for a bit of a fluorescent green towards a darker green which we apply or the blue of the lake also in a darker shade. This allows to manipulate in a global way the color palette of this map. Finally, the information on the position of the image can be found either in the header of a tif file and it is called GeoTiff, or in a word file that accompanies the image file with the tfw, JGW extensions etc.

Notes

Summary



19m 11s