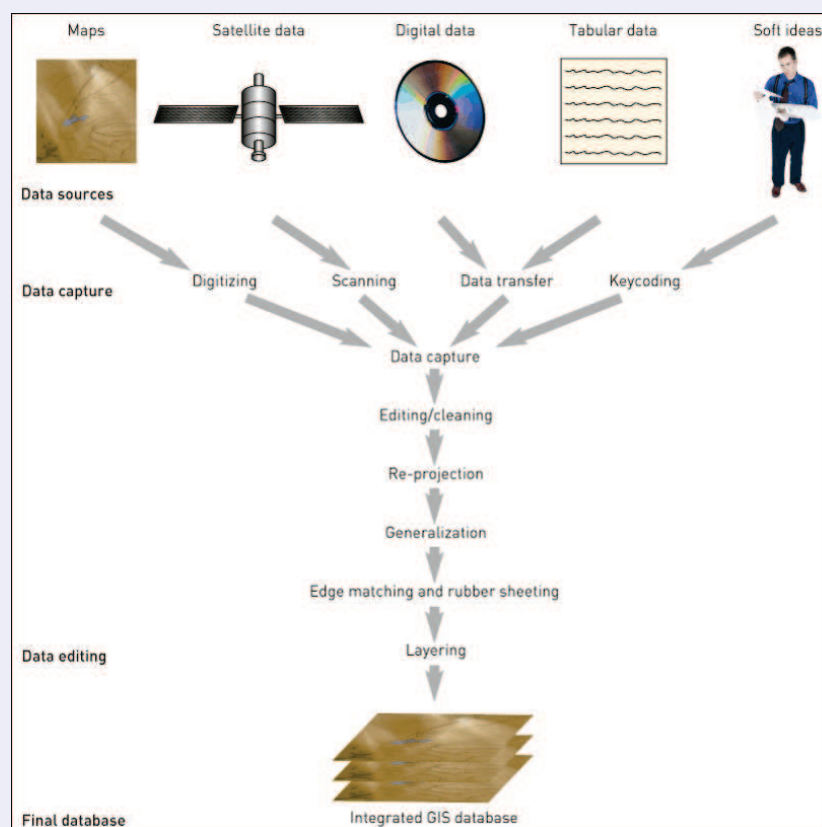


de *Información Geográfica*

José Samos Jiménez

2020 jsamos (lsi-ugr)
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Granada

Curso 2020-21



Transformación e Integración

- Detectar y corregir errores:
 - ▶ Datos con origen común (errores ocultos) vs. datos de distinto origen (errores evidentes).
- Resolver heterogeneidades:
 - ▶ Técnicas, del modelo de datos y semánticas.
- *Concatenación*:
 - ▶ Integración de dos o más fuentes de datos.
 - ▶ Conjunto de funciones para superar las diferencias entre los conjuntos de datos o fusionar su contenido.
 - ★ Reemplazar dos o más versiones de la misma información por una que refleje su agrupación.
 - ★ Expresar de alguna forma la incertidumbre que incluye el resultado.
 - ▶ Uso de Metadatos.

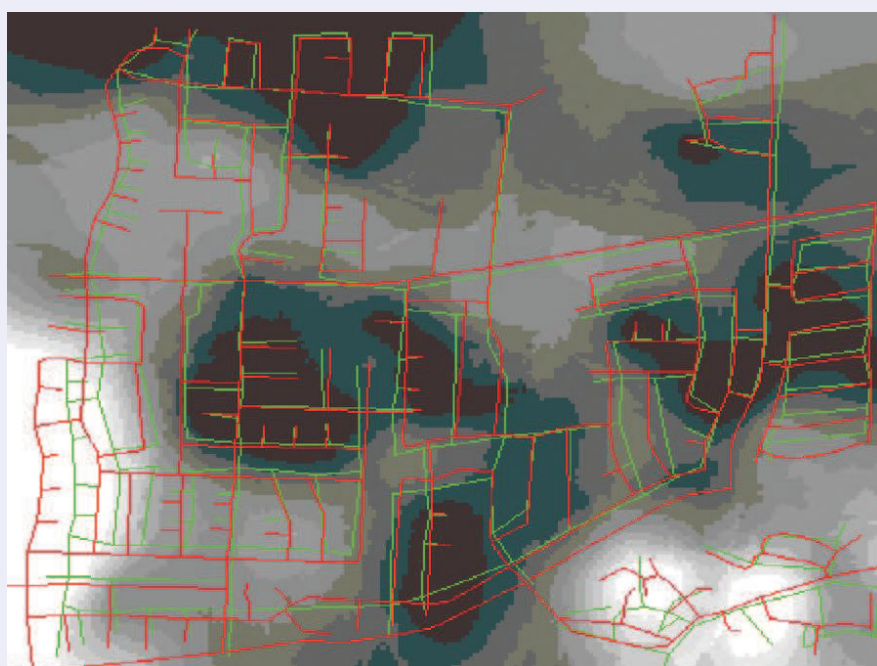
Contenido

- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
- 3 Metadatos
- 4 Herramientas
- 5 Conclusiones
- 6 Bibliografía

Detectar y corregir errores

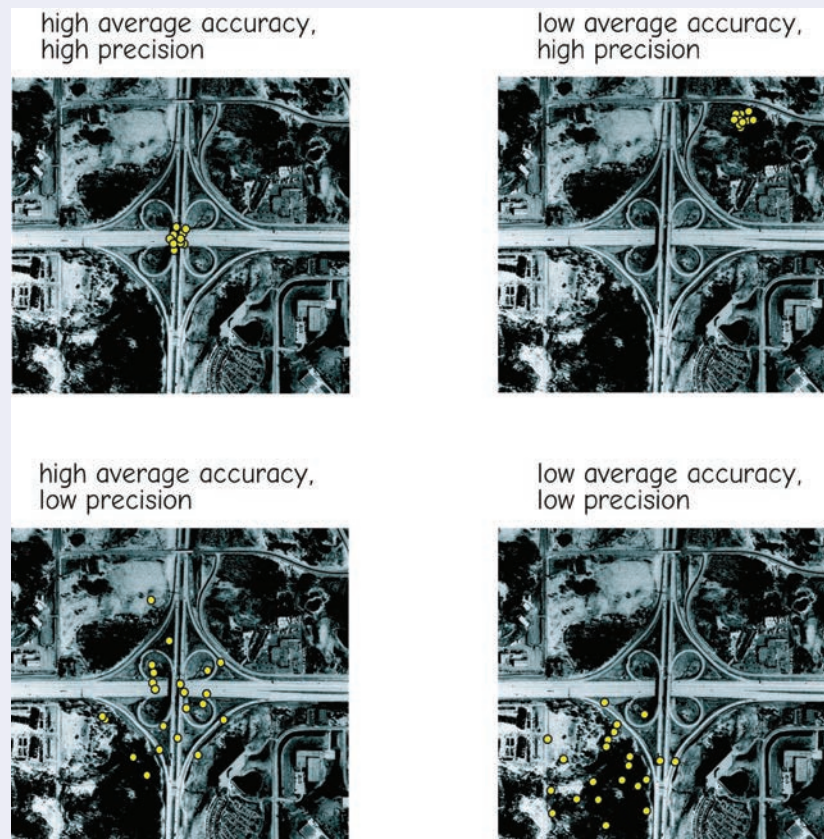
- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
- 3 Metadatos
- 4 Herramientas
- 5 Conclusiones
- 6 Bibliografía

Ejemplo: superposición de datos de distinto origen

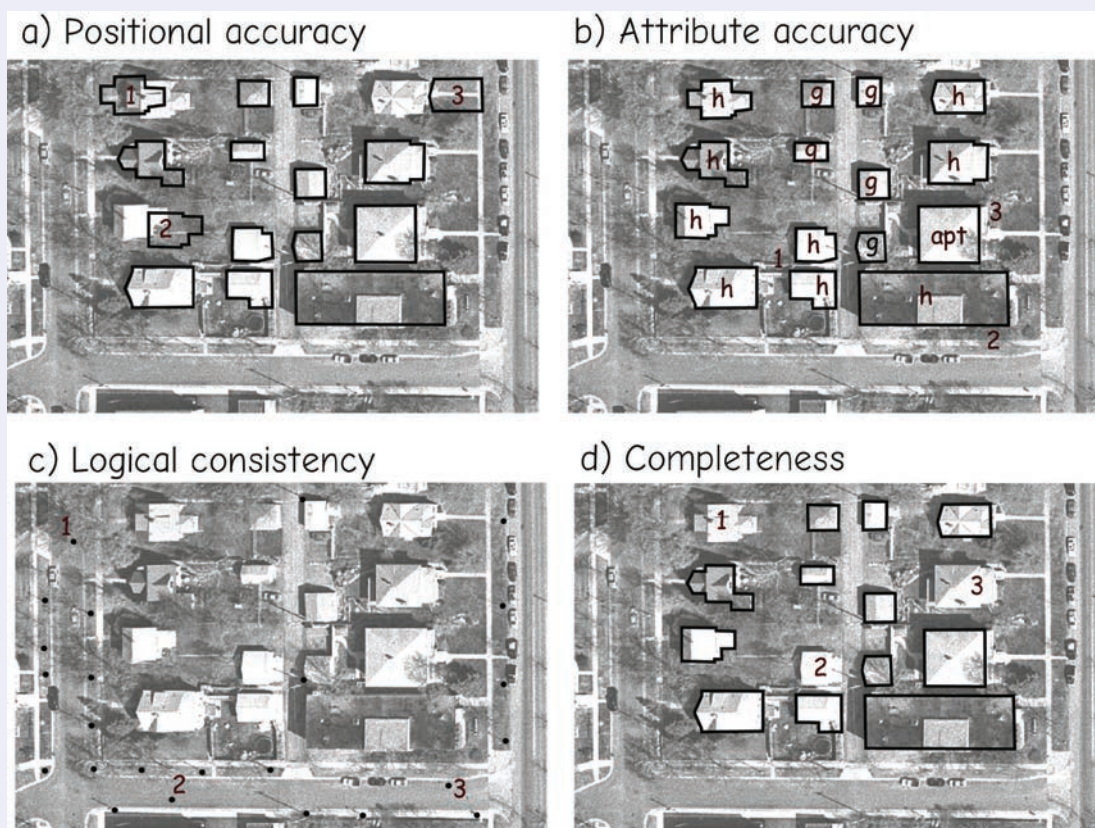


- Calles generadas por dos empresas distintas.
- Hay fallos de más de 100 m.

Exactitud, precisión y sesgo



Ejemplo: errores de varios tipos (numerados)



Tipos de errores

- Errores en las coordenadas.
- Datos duplicados.
- Datos faltantes.
- Imprecisión o incertidumbre en atributos.

Origen de los errores

- Derivados de la fuente de datos original.
 - ▶ Difíciles de identificar.
- Introducidos durante el proceso de transformación.
 - ▶ Al cambiar los datos de formato.

Errores en las coordenadas

Errores de escala y sistema de coordenadas

- Datos con distinto origen \Rightarrow Definir un origen común.
- Diferentes unidades de medida \Rightarrow Considerar la más gruesa.
 - ▶ El resultado puede ser tan bueno como el peor.
 - ★ Ejemplo: unos datos en m. y otros en 10 m.
- Diferente orientación \Rightarrow Rotación.
- Confusión en coordenadas geográficas: (*latitud, longitud*) es (y, x).



Resolver heterogeneidades

- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
 - Heterogeneidad técnica
 - Heterogeneidad del modelo de datos
- 3 Metadatos
- 4 Herramientas
- 5 Conclusiones
- 6 Bibliografía

Tipos de heterogeneidades en los datos

Heterogeneidad técnica

- Diferente formato en los datos.
 - ▶ ⇒ *Transformación*: conversión de formato, conversión del sistema de coordenadas.

Heterogeneidad del modelo de datos

- Diferentes maneras de representar y almacenar los mismos datos.
 - ▶ ⇒ *Transformación*.

Heterogeneidad semántica

- Distinto significado de los datos.
 - ▶ ⇒ *Transformación*.

Para conseguir la *integración* de los datos.

Heterogeneidad técnica

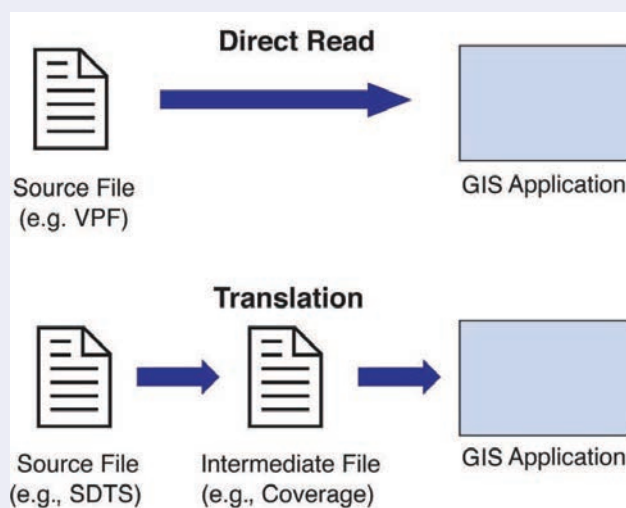
Formatos de uso frecuente para datos espaciales (i)

Type and source	Extension	Characteristics (R=Raster, V=Vector, A=Attribute, I=Image)
Comma Separated Value	.csv	Common ASCII text format used to distribute attribute and often vector information (A, V).
DXF, AutoDesk	.dxf	Drawing exchange file, an ASCII or binary file for exchanging spatial data (V)
DWG, Autodesk	.dwg	Native binary file used by AutoDesk to store geographic data and drawings in AutoCAD (V)
Geodatabase, ESRI	.gdb, .mdb	ESRI container for many data types (R, V, A, I)
GeoJSON, open standard	.json, .geojson	Open standard for representing and displaying simple geographic features (V, A).
GeoPackage, open standard	.gpkg	Open standard for representing vector and raster data, compatible with SQLite
GeoTIFF, open standard	.TIF, .TFF	An extension for georeferencing Aldus-Adobe public domain TIFF format (R)
GPX, open standard	.gpx	A specification based on XML for basic GNSS data
Imagine, ERDAS	.img	Multiband capable image format (R)

Formatos de uso frecuente para datos espaciales (y ii)

Type and source	Extension	Characteristics (R=Raster, V=Vector, A=Attribute, I=Image)
Interchange, ESRI	.e00	ASCII text file for vector and identifying attribute data (V).
Keyhole Markup Language, Google	.KML	XML extension for displaying and annotating features and images (V, I, A)
LAS, ASPRS	.LAS	Laser point cloud data storage (V)
shapefile, ESRI	.shp, .shx, .dbf, .prj, and others	Three or more binary files that include the vector coordinate, attribute, and other information (V)
TIGER, U.S. Census	tgrxyyy, stfzz	Set of files by U.S. census areas, xx is a state code, yyy an area code, zz numbers for various file types
MIF/MID, MapInfo	.mif, .mid	Map Interchange File, vector and raster data transport from MapInfo (V,R)
NetCDF, OGC	.cdf	Machine-independent data formats for scientific data arrays (R, A, I)
NLAPS, NASA	various in a directory	Image data from various Landsat satellites, in a specified directory structure (I, R)
SDTS, U.S. Government	none	Spatial Data Transfer Standard, specifies the spatial objects, attributes, reference system (R,V, A)

Importación directa o conversión previa

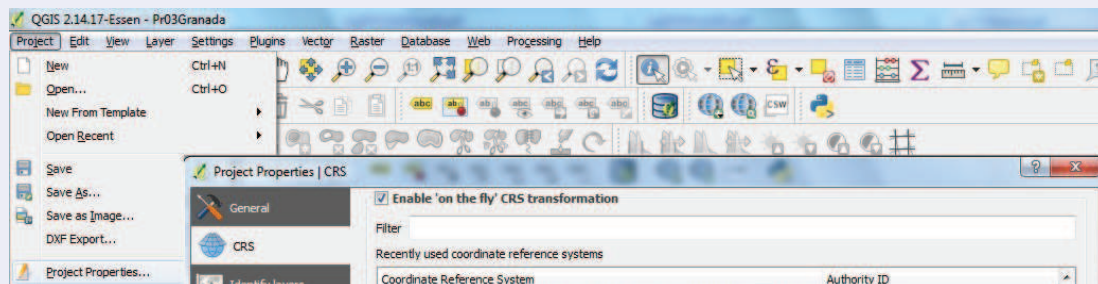


- Ejemplos de conversores:

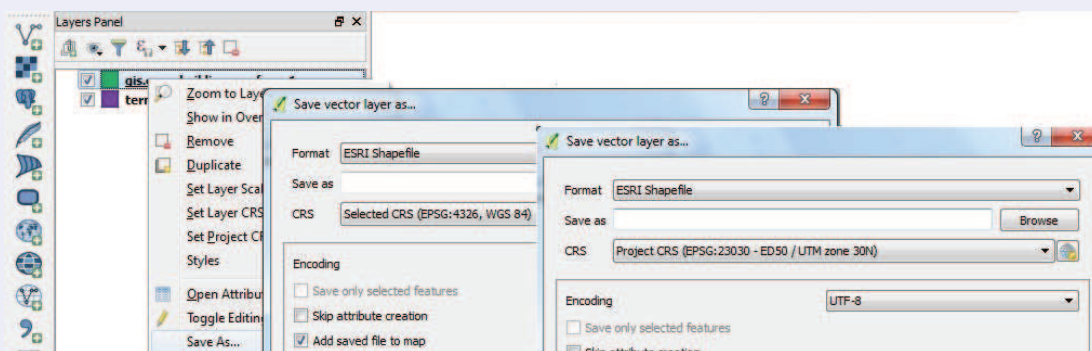
- ▶ *Geospatial Power Tools: GDAL/OGR*
- ▶ <https://www.gisconvert.com/>
- ▶ <https://mygeodata.cloud/converter/>

Capas con distinto CRS

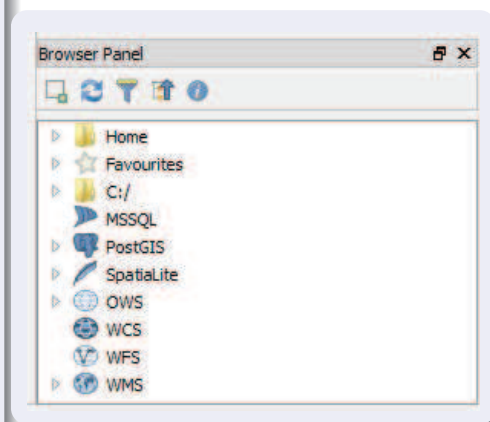
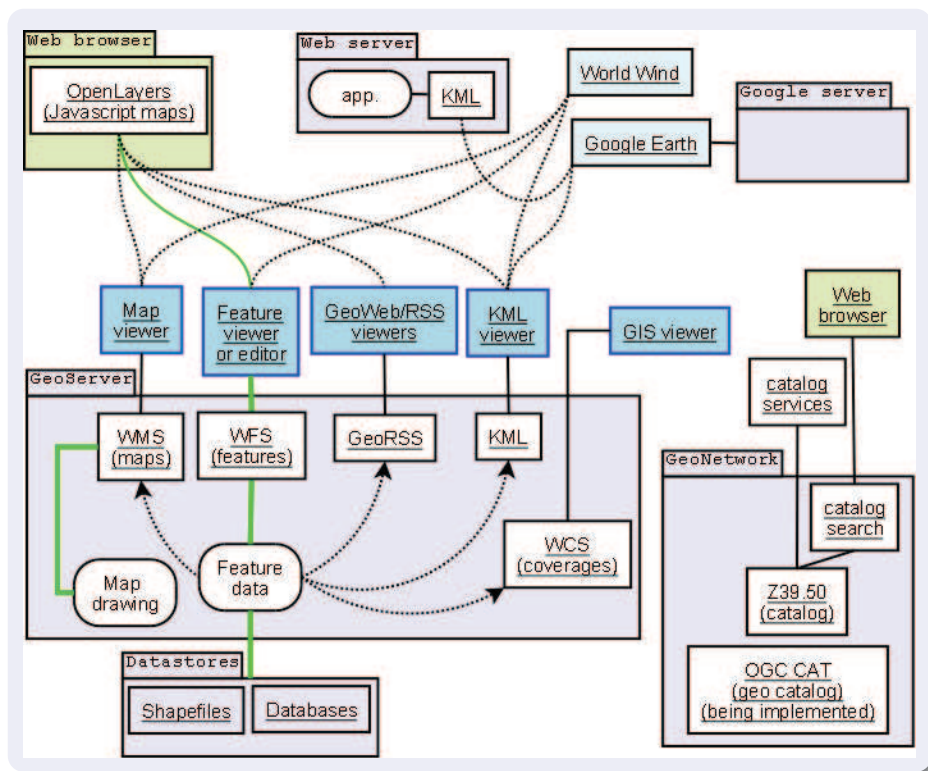
Enable "on the fly" CRS transformation



Conversión para poder hacer operaciones



Formatos estándar para acceso a los datos (i)



Formatos estándar para acceso a los datos (y ii)

OWS

OGC Web Services.

WCS

Web Coverage Service: acceso a datos de cobertura (objetos que cubren un área geográfica).

WFS

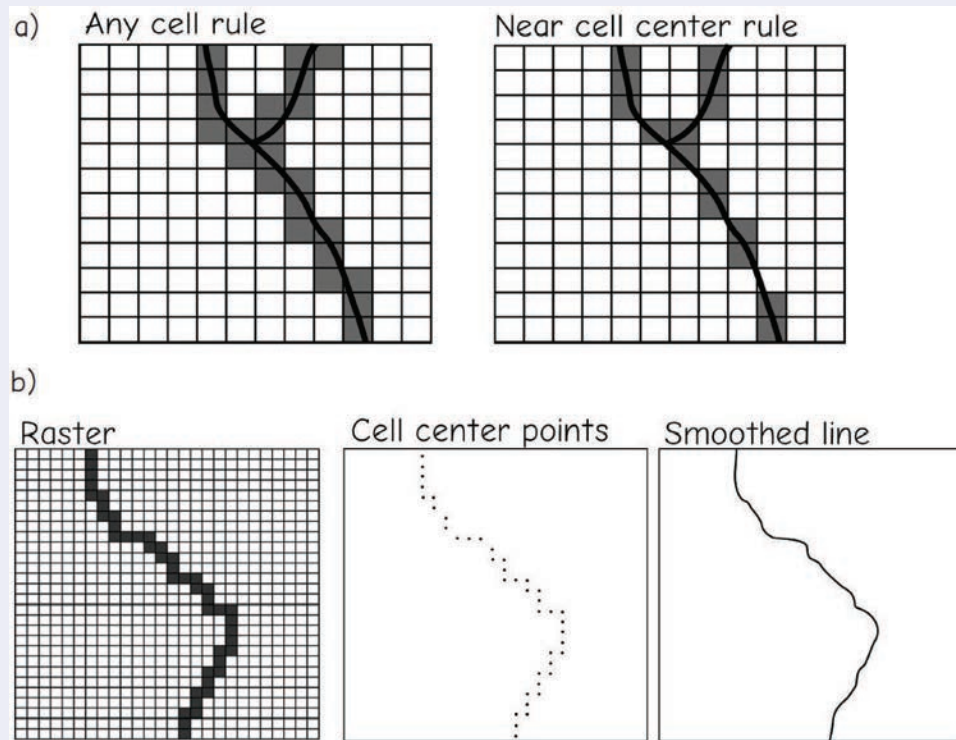
Web Feature Service: operaciones para descubrir, consultar y transformar datos geográficos.

WMS

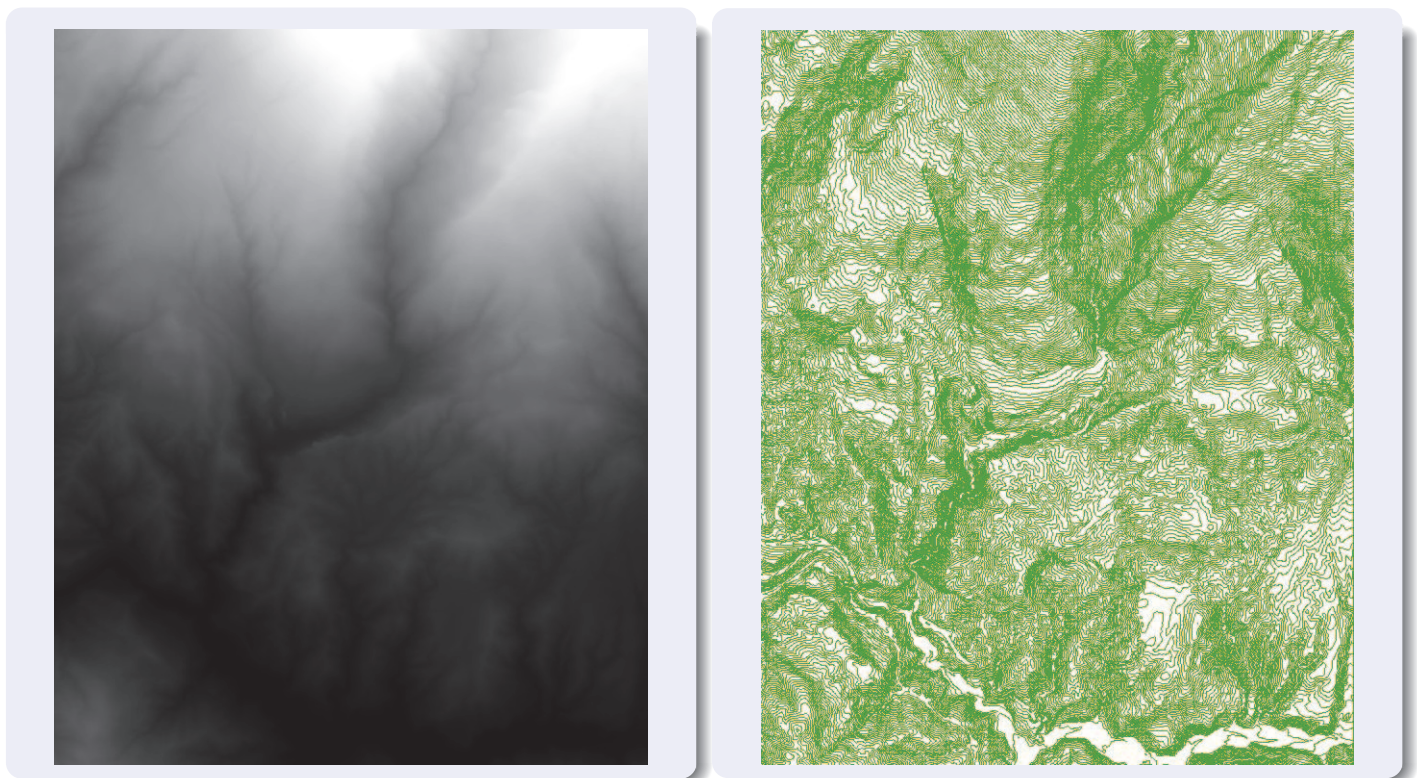
Web Map Service: obtención de imágenes geo-registradas.

Heterogeneidad del modelo de datos

Conversión entre Ráster y Vectorial



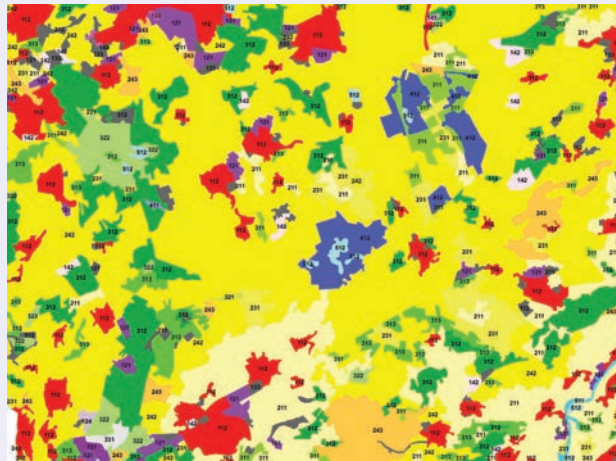
Obtención de curvas de nivel



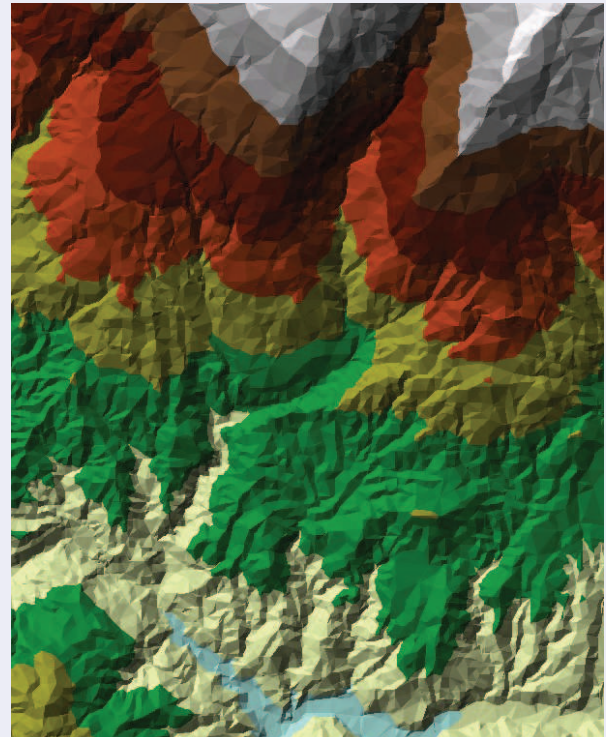
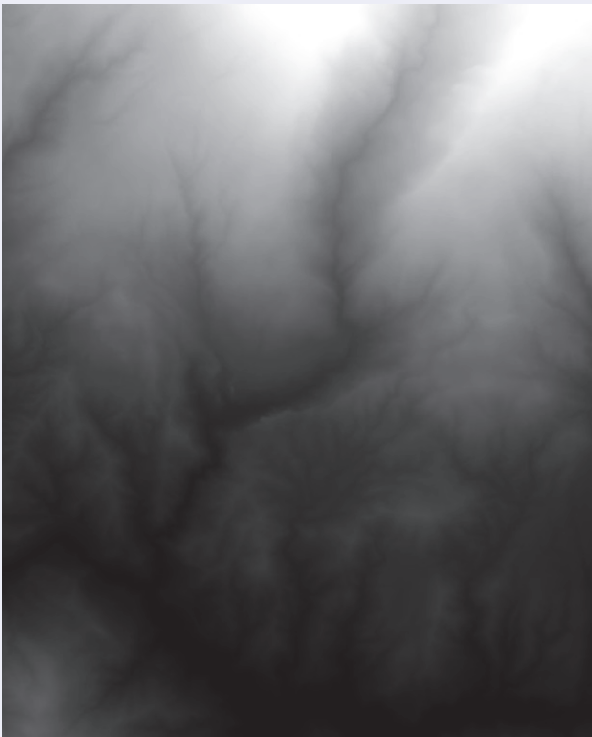
Otras operaciones sobre ráster

- Modificar la resolución
- Fusionar
- Recortar
- Clasificar en niveles

Transformar en polígonos

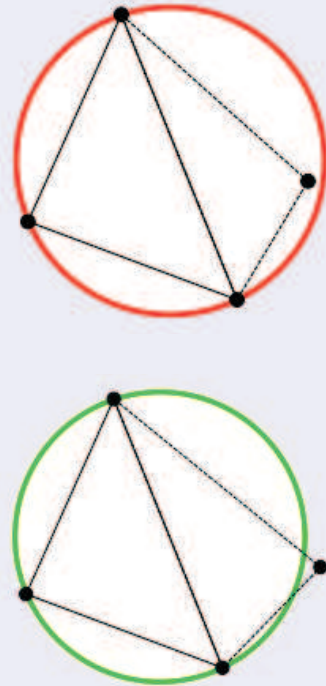


Obtención de TIN



Creación de TIN

Triangulación de Delaunay



https://es.wikipedia.org/wiki/Triangulaci%C3%B3n_de_Delaunay

Diagrama de Voronoi

Triangulación de Delaunay



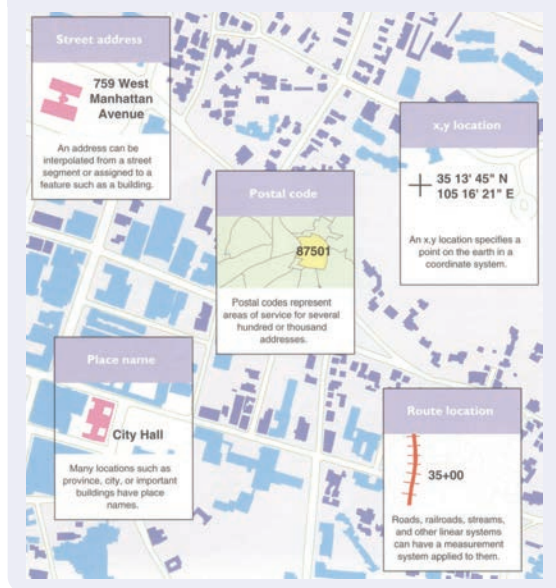
Teselación de Voronoi o Polígonos de Thiessen



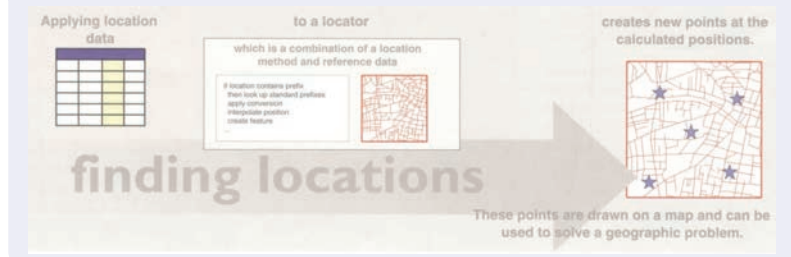
Centros de las circunferencias.

Encontrar ubicaciones

Tipos de ubicaciones



Proceso de localización



Metadatos

- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
- 3 Metadatos**
- 4 Herramientas
- 5 Conclusiones
- 6 Bibliografía

Información que deben contener los metadatos geográficos

- ¿Qué describe el conjunto de datos?
 - ▶ Título, área geográfica, tiempo, tipo de datos.
- ¿Quién lo produjo?
 - ▶ Autor y administrador.
- ¿Por qué fue creado?
 - ▶ Objetivos del proyecto y del conjunto de datos, cómo usarlo.
- ¿Cómo fue creado?
 - ▶ De dónde procede, cómo se generó, otros datos relacionados.
- ¿Cómo son de fiables los datos? ¿Qué problemas presenta?
- ¿Cómo se puede obtener una copia?
 - ▶ Restricciones, formatos.
- ¿Quién generó los metadatos? ¿Qué estándar siguen?

Estándares de metadatos

Proliferación de estándares: de organismos y de países.

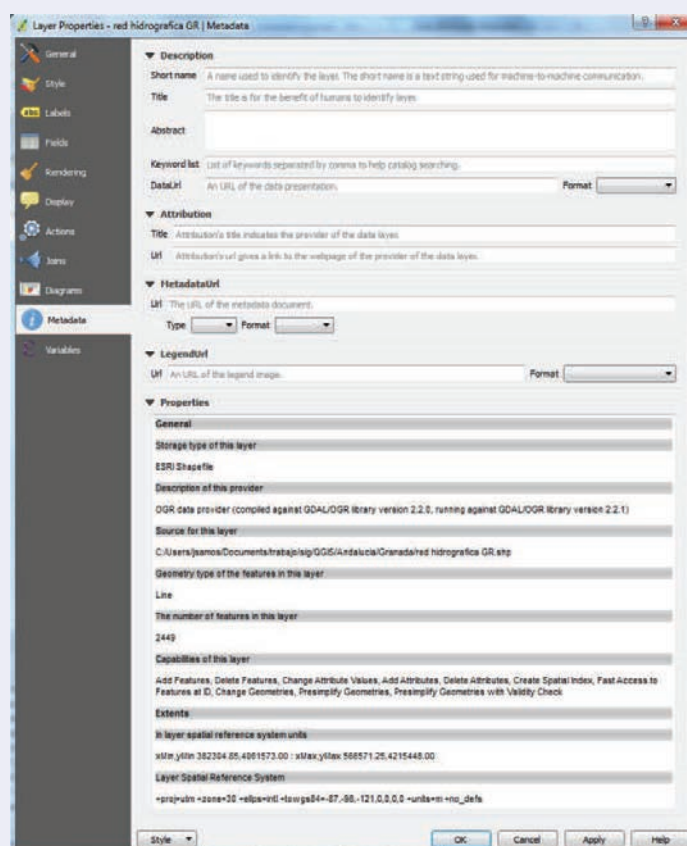
- *OGC Catalogue Services 3.0 - General Model:*
 - ▶ <http://docs.openeospatial.org/is/12-168r6/12-168r6.html>
- *Federal Geographic Data Committee - Content Standard for Digital Geospatial Metadata:*
 - ▶ <https://www.fgdc.gov/metadata/csdgm/>
- *ISO 19115-1:2014 - Geographic information - Metadata:*
 - ▶ <https://www.iso.org/standard/53798.html>
 - ▶ FGDC ha sido miembro activo, su estándar ha tenido mucho impacto.
 - ▶ Cada país puede definir su propio perfil.
- *Dublin Core Metadata Standard:*
 - ▶ <http://dublincore.org/documents/dces/>

Formato de los metadatos

Los estándares no imponen el formato.

- Texto plano.
- Esquema (jerárquico).
- FAQ (*Frequently Asked Questions*).
- SGML.
- HTML.
- XML.

Ejemplo de metadatos (QGIS)



Ejemplo de metadatos (MIT)

<https://geodata.mit.edu/catalog/princeton-z603qz96f>

Institution

MIT

Data Type

Line

Theme Keywords

line rivers drainage systems hydrography

Place Keywords

World

Base Layers -
Log In

Download FGDC Metadata

- Identification Information
- Metadata Reference Information
- Spatial Data Organization Information
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Data Quality Information

```

1 <metadata xml:lang="en">
2   <Esrri>
3     <idinfo>
4       <metainfo>
5         <spdoinfo>
6           <sprf>
7             <esinfo>
8               <distinfo>
9                 <datequal>
10                  <dataIdInfo>
11                    <mdLang>
12                      <cmdStanName Sync="TRUE">ISO 19115 Geographic Information - Metadata</mdStanName>
13                      <cmdStanVer Sync="TRUE">DIS_ESRI.0</mdStanVer>
14                      <mdChar>
15                        <mdHrLv>
16                          <mdHrLvName Sync="TRUE">dataset</mdHrLvName>
17                          <distInfo>
18                            <refSysInfo>
19                              <spatRepInfo>
20                                <mdDateSt Sync="TRUE">20090127</mdDateSt>
21                                </mdDateSt>
22                              </refSysInfo>
23                            </distInfo>
24                          </mdHrLv>
25                        </mdChar>
26                      </cmdStanVer>
27                    </mdLang>
28                  </dataIdInfo>
29                </datequal>
30              </distinfo>
31            </esinfo>
32          </sprf>
33        </spdoinfo>
34      </metainfo>
35    </idinfo>
36  </Esrri>
37 </metadata>
```

Ejemplo de metadatos

roadtr|020.txt

Identification Information:

Citation:

Citation_Information:

Originator: U.S. Geological Survey

Publication Date: 199911

Title: Major Roads of the United States

Publication Information:

Publication Place: Reston, Virginia

Publisher: U.S. Geological Survey

Online_Linkage: <<http://nationalatlas.gov/atlasftp.html>>

Description:

Abstract:

This data set portrays the major roads in the United States, Puerto Rico, and the U.S. Virgin Islands. The file was produced by joining the individual State roads layers from the 1:2,000,000-scale Digital Line Graph (DLG) data produced by the USGS. This is a revised version of the March 1999 data set.

Purpose:

These data are intended for geographic display and analysis at the national level, and for large regional areas. The data should be displayed and analyzed at scales appropriate for 1:2,000,000-scale data. No responsibility is assumed by the U.S. Geological Survey in the use of these data.

Time_Period_of_Content:

Time_Period_Information:

Range_of_Dates/Times:

```
Beginning_Date: 1972
```

Ending_Date: 1999

Currentness_Reference: ground condition

...

Herramientas

- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
- 3 Metadatos
- 4 Herramientas**
 - Geospatial Power Tools
 - GeoKettle
- 5 Conclusiones
- 6 Bibliografía

Geospatial Power Tools

<https://www.openhub.net/p/gdal>

BLACKDUCK | Open Hub

Projects People Organizations Tools Blog

GDAL Settings Report Duplicate

Follow OH SIGN IN JOIN NOW

Projects Search...

GDAL 133 Very High Activity 1 Use This!

Analyzed 3 days ago, based on code collected 8 months ago.

Project Summary

GDAL is a translator library for raster geospatial data formats. The related OGR library (which lives within the GDAL source tree) provides a similar capability for simple features vector data.

Tags

gdal geospatial gis mapping ogr osgeo

In a Nutshell, GDAL...

- ... has had 62,440 commits made by 187 contributors representing 2,058,761 lines of code
- ... is mostly written in C++ with an average number of source code comments
- ... has a well established, mature codebase maintained by a very large development team with decreasing Y-O-Y commits
- ... took an estimated 594 years of effort (COCOMO model) starting with its first commit in October, 1998 ending with its most recent commit 3 days ago

Quick Reference

Organization: OSGeo - The Open Source Geospatial Foundation

Project Links: [Homepage](#) [Documentation](#) [Download](#) [Forums](#) [Issue Trackers](#)

Code Locations: (2 Locations)

Similar Projects: [gvSIG Desktop 1](#) [spatialguru](#) [cl-gdal](#) [Fiona \(Python\)](#)

Managers: Howard Butler

GDAL (*Geospatial Data Abstraction Library*)

Ejemplo (i)

Formatos ráster soportados

```
C:\>gdalinfo --formats
Supported Formats:
  ECW -raster- (rov): ERDAS Compressed Wavelets (SDK 5.3)
  JP2ECW -raster,vector- (rov): ERDAS JPEG2000 (SDK 5.3)
  VRT -raster- (rw+v): Virtual Raster
  DERIVED -raster- (ro): Derived datasets using VRT pixel functions
  GTiff -raster- (rw+vs): GeoTIFF
  NITF -raster- (rw+vs): National Imagery Transmission Format
  RPFTOC -raster- (rovs): Raster Product Format TOC format
  ECRGTOC -raster- (rovs): ECRG TOC format
  HFA -raster- (rw+v): Erdas Imagine Images (.img)
  SAR_CEOS -raster- (rov): CEOS SAR Image
  CEOS -raster- (rov): CEOS Image
  JAXAPALSAR -raster- (rov): JAXA PALSAR Product Reader (Level 1.1/1.5)
  GFF -raster- (rov): Ground-based SAR Applications Testbed File Format (.gff)
  ELAS -raster- (rw+v): ELAS
  AIG -raster- (rov): Arc/Info Binary Grid
  AAIGrid -raster- (rwv): Arc/Info ASCII Grid
  GRASSASCIIGrid -raster- (rov): GRASS ASCII Grid
  SDTS -raster- (rov): SDTS Raster
  OGD1 -raster- (ros): OGD1 Bridge
  DTED -raster- (rwv): DTED Elevation Raster
  PNG -raster- (rwv): Portable Network Graphics
  JPEG -raster- (rwv): JPEG JFIF
  MEM -raster- (rw+): In Memory Raster
  JDEM -raster- (rov): Japanese DEM (.mem)
  GIF -raster- (rwv): Graphics Interchange Format (.gif)
  BIGGIF -raster- (rov): Graphics Interchange Format (.gif)
  ...
```

Ejemplo (ii)

Información estadística

```
C:\datos>gdalinfo -stats h10_1042_1-2.tif
Driver: GTiff/GeoTIFF
Files: h10_1042_1-2.tif
Size is 767, 488
Coordinate System is:
PROJCS["ETRS89 / UTM zone 30N",
...
  AUTHORITY["EPSG","25830"]]
Origin = (453495.0000000000000000,4090485.0000000000000000)
Pixel Size = (10.000000000000000,-10.000000000000000)
Metadata:
  AREA_OR_POINT=Point
Image Structure Metadata:
  INTERLEAVE=BAND
Corner Coordinates:
Upper Left  ( 453495.000, 4090485.000) ( 3d31'20.60"W, 36d57'33.48"N)
Lower Left  ( 453495.000, 4085605.000) ( 3d31'19.52"W, 36d54'55.12"N)
Upper Right ( 461165.000, 4090485.000) ( 3d26'10.44"W, 36d57'34.74"N)
Lower Right ( 461165.000, 4085605.000) ( 3d26' 9.54"W, 36d54'56.37"N)
Center      ( 457330.000, 4088045.000) ( 3d28'45.03"W, 36d56'14.96"N)
Band 1 Block=767x1 Type=Float32, ColorInterp=Gray
  Minimum=506.982, Maximum=2033.411, Mean=1181.350, StdDev=347.595
  Unit Type: metre
Metadata:
  STATISTICS_MAXIMUM=2033.4110107422
  STATISTICS_MEAN=1181.3501471428
  STATISTICS_MINIMUM=506.98199462891
  STATISTICS_STDDEV=347.59519212124
```

Ejemplo (iii)

Convertir un ráster

```
C:\datos>gdal_translate -of GTiff -co "TILED=YES" h10_1042_1-2.tif out_tiled.tif
Input file size is 767, 488
0...10...20...30...40...50...60...70...80...90...100 - done.
```

Pasa de una tira unidimensional a «tiles» bidimensionales.

- Cada «tile» es del mismo tamaño y se comprime independientemente.

```
C:\datos>gdal_translate -outsize 20% 20% h10_1042_1-2.tif out.tif
Input file size is 767, 488
0...10...20...30...40...50...60...70...80...90...100 - done.
```

Reduce el tamaño al 20 % de la entrada.

```
C:\datos>gdalwarp -t_srs "epsg:3035" h10_1042_1-2.tif out2.tif
0...10...20...30...40...50...60...70...80...90...100 - done.
```

Reproyectar.

Ejemplo (y iv)

De ráster a vectorial

```
C:\datos>gdal_contour -a elev h10_1042_1-2.tif contour.shp -i 10.0
0...10...20...30...40...50...60...70...80...90...100 - done.
```

OGR (*OpenGIS Simple Features Reference Implementation*)

Ejemplo (i)

Formatos vectoriales soportados

```
C:\>ogrinfo --formats
Supported Formats:
JP2ECW -raster,vector- (rov): ERDAS JPEG2000 (SDK 5.3)
PCIDSK -raster,vector- (rw+v): PCIDSK Database File
netCDF -raster,vector- (rw+s): Network Common Data Format
JP2OpenJPEG -raster,vector- (rwv): JPEG-2000 driver based on OpenJPEG library
PDF -raster,vector- (rw+vs): Geospatial PDF
DB2ODBC -raster,vector- (rw+): IBM DB2 Spatial Database
ESRI Shapefile -vector- (rw+v): ESRI Shapefile
MapInfo File -vector- (rw+v): MapInfo File
UK .NTF -vector- (ro): UK .NTF
OGR_SDTs -vector- (ro): SDTS
S57 -vector- (rw+v): IHO S-57 (ENC)
DGN -vector- (rw+): Microstation DGN
OGR_VRT -vector- (rov): VRT - Virtual Datasource
REC -vector- (ro): EPIInfo .REC
Memory -vector- (rw+): Memory
BNA -vector- (rw+v): Atlas BNA
CSV -vector- (rw+v): Comma Separated Value (.csv)
NAS -vector- (ro): NAS - ALKIS
GML -vector- (rw+v): Geography Markup Language (GML)
GPX -vector- (rw+v): GPX
LIBKML -vector- (rw+v): Keyhole Markup Language (LIBKML)
KML -vector- (rw+v): Keyhole Markup Language (KML)
GeoJSON -vector- (rw+v): GeoJSON
OGR_GMT -vector- (rw+): GMT ASCII Vectors (.gmt)
GPKG -raster,vector- (rw+vs): GeoPackage
SQLite -vector- (rw+v): SQLite / Spatialite
ODBC -vector- (rw+): ODBC
...
```


Ejemplo (ii)

Información de un archivo

```
C:\datos>ogrinfo -so -al shp9108647358650951425.shp
INFO: Open of 'shp9108647358650951425.shp'
      using driver 'ESRI Shapefile' successful.
Layer name: shp9108647358650951425
Metadata:
  DBF_DATE_LAST_UPDATE=1918-08-02
Geometry: Line String
Feature Count: 16696
Extent: (100884.737000, 3987866.500000) - (621194.657700, 4288459.000000)
Layer SRS WKT:
PROJCS["ED50 / UTM zone 30N",
...
AUTHORITY["EPSG","23030"]]
matricula: String (254.0)
...
shape_leng: Real (33.31)
cod_ent: String (254.0)

C:\datos>ogrinfo -so -al restaurants.csv
INFO: Open of 'restaurants.csv'
      using driver 'CSV' successful.
Layer name: restaurants
Geometry: None
Feature Count: 50002
Layer SRS WKT:
(unknown)
field_1: String (0.0)
field_2: String (0.0)
field_3: String (0.0)
```

Ejemplo (y iii)

Convertir un vector

```
C:\datos>ogr2ogr -f KML out.kml pantanos_GR.shp

C:\datos>ogr2ogr -f GML out.gml pantanos_GR.shp

C:\datos>ogr2ogr -t_srs "epsg:3035" out.shp pantanos_GR.shp
```

PRJ.4

PRJ.4: utilidades de proyección

- proj e invproj
- cs2cs
- gdalsrsinfo

GDAL/OGR desde R

<https://CRAN.R-project.org/package=rgdal>

- Provides bindings to the 'Geospatial' Data Abstraction Library ('GDAL') ($\geq 1.11.4$) and access to projection/transformation operations from the 'PROJ.4' library.
- The 'GDAL' and 'PROJ.4' libraries are external to the package, and, when installing the package from source, must be correctly installed first.
- Más información:
<https://cran.r-project.org/web/packages/rgdal/rgdal.pdf>

GeoKettle

GeoKettle

<http://www.spatialytics.org/projects/geokettle/>



- Spatial ETL tool dedicated to the integration of different spatial data sources for building and updating geospatial data warehouses.

Extract data from

- Spatial database types: PostGIS, Oracle spatial, MySQL, Microsoft SQL Server 2008, Ingres and IBM DB2
- SOLAP (Spatial OLAP) system: GeoMondrian
- Geo files (data formats): Shapefile, GML, KML, OGR
- OGC Web services: Sensor Observation Service (SOS), Catalogue Web Service (CSW)

Transformation

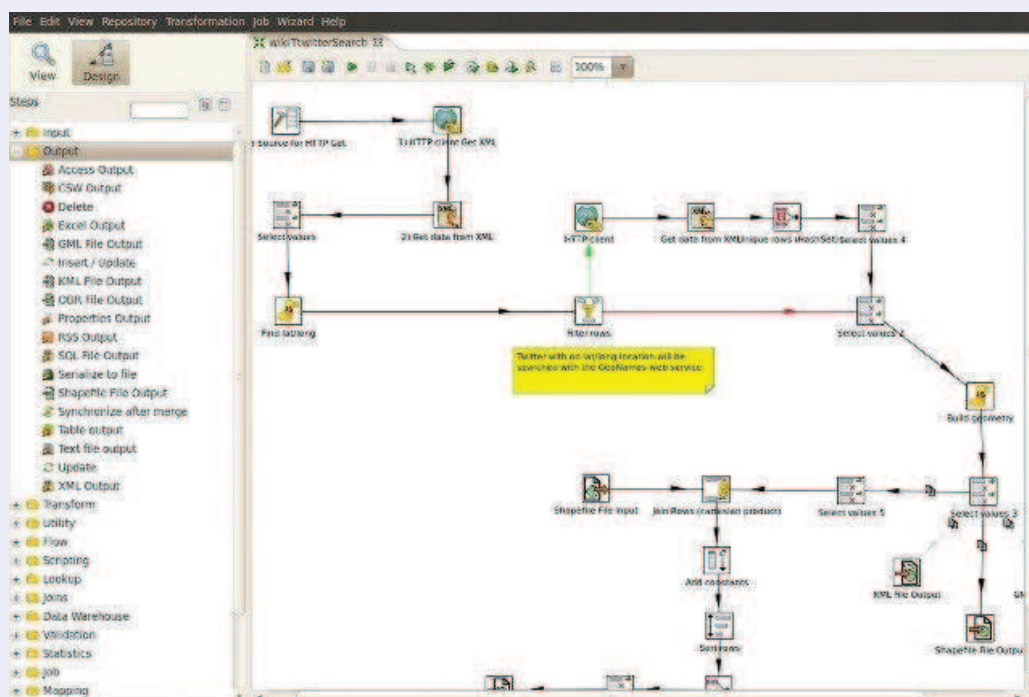
Transformation of data

- Calculating:
 - ▶ Buffers, Centroid, Random point on surface, Area, Length, Distance, Intersection, Union, Envelope, Boundary, Convex hull, Difference, Symetric difference, Inverse geometry
- Geoprocessing:
 - ▶ General: Clip, Clip with rectangle, Split multipart
 - ▶ Points: Delaunay algorithm
 - ▶ Lines: Polylines → polygons, Simplify lines, Smooth lines, Polylines to single segments, Split polylines at nodes, Split polylines with points
 - ▶ Polygons: Simplify polygons, Remove holes, Polygons → polylines

Load data into a target format

- Spatial database loads
- Spatial data warehouse population
- Data formats: Shapefile, GML, KML, OGR
- OGC Web services: Catalogue Web Service (CSW)

Ejemplo



Conclusiones

- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
- 3 Metadatos
- 4 Herramientas
- 5 Conclusiones**
- 6 Bibliografía

Conclusiones

- El poder de los SIG está en combinar o integrar datos de varios tipos y analizarlos conjuntamente.
- Usaremos datos de distintas fuentes.
- Todas las capas deben usar el mismo Sistema de Coordenadas.
- El uso de estándares facilita el intercambio de datos (heterogeneidad técnica).
- Realizaremos transformaciones para resolver:
 - ▶ Heterogeneidad del modelo de datos.
 - ▶ Heterogeneidad semántica.
- La calidad del resultado puede ser tan buena como la del peor conjunto usado.
- Los metadatos favorecen el proceso de integración de datos.

Bibliografía

- 1 Detectar y corregir errores
- 2 Resolver heterogeneidades
- 3 Metadatos
- 4 Herramientas
- 5 Conclusiones
- 6 Bibliografía

Bibliografía

- Bha11** Basudeb Bhatta. *Remote Sensing and GIS (Second Edition)*. Oxford, 2011.
- Bol16** Paul Bolstad. *GIS Fundamentals (Fifth Edition)*. XanEdu, 2016.
- HCC11** Ian Heywood, Sarah Cornelius, and Steve Carver. *An Introduction to Geographical Information Systems (Fourth Edition)*. Pearson, 2011.
- LGMR15** Paul A. Longley, Michel F. Goodchild, David J. Maguire, and David W. Rhind. *Geographic Information Science and Systems (Fourth Edition)*. Wiley, 2015.
- Zei99** Michael Zeiler. *Modeling Our World*. ESRI Press, 1999.