Automate your GIS – Scripting in Python
5 credits (hp)

Fredrik Lindberg

Urban Climate Group
Department of Earth Sciences
University of Gothenburg

Course outline

Course focusing on hands-on GIS

1. First week – Practical tutorials and exercises
2. Second and third week – Project work

Python programming
Python and GIS
Programming and scripting using QGIS
QGIS (processing) plugin building
Tools for spatial manipulation/processing
ArcGIS and Python

Project work

Preferable something related to your PhD-work

For example:

- Big data handling
- Automate GIS processing/analysis
- Build a plugin for your spatial model
- Tryout and demonstrate certain GIS-technique using command line and scripting
- Make a tutorial based on some spatial library available

Hand in the form of a written report including your code and results

Presentation of your work in the final seminar

Talk today

- Cocktail talk (participants)
- Geodata formats
- Coordinate reference systems
- Tools and libraries
- Command line execution
- Scripting
- Open source GIS

Proprietary formats and software

Propriatary data formats

Many software companies have their own “closed” format.

Software	Data format
ESRI	Geodatabase, ERSI grid
Autodesk	.dwg (CAD)
Rihno3D	.dwg (CAD)
ERDAS IMAGINE	.img
MapInfo	.tab
FME Safe (ETL)	In house (format independent)

Data formats

OS GIS data formats

As OS software need protection (by licensing) in order to be considered as open source, this is also needed for data formats.

Open format with regards to data means that **they have a published specification** you can use to **write applications and utilities** that work with the format. A **closed format requires** you to use the **vendor-provided API**. Of course, this is only a concern if you want to write an application of your own.

Although it is not important to understand the formats to use them, it does help to have basic knowledge about different formats.

An application programming interface (**API**) is a particular set of rules and specifications that software programs can follow to communicate with each other

Data formats

OS GIS data formats (cont.)

Since most OSGIS software use GDAL/OGR libraries where most GIS formats are included for reading and writing, formats are usually not a problem.

The need for standards

Because of the many different formats and owners of formats there is a **need for standardization**. When it comes to open source GIS data formats, this is done by Open Geospatial Consortium (OGC).

The need for standards has **increased with increased number of users** (on line and off line).

Dealing with standards is **very boring** (own opinion). An example; for the GML-format (upcoming slides) a document of 436 pages exists specifying how to set up this particular format.

Data formats

Repetition: Raster - Vector

Raster

Each cell (pixel) contains one value

DEM – Digital Elevation Model

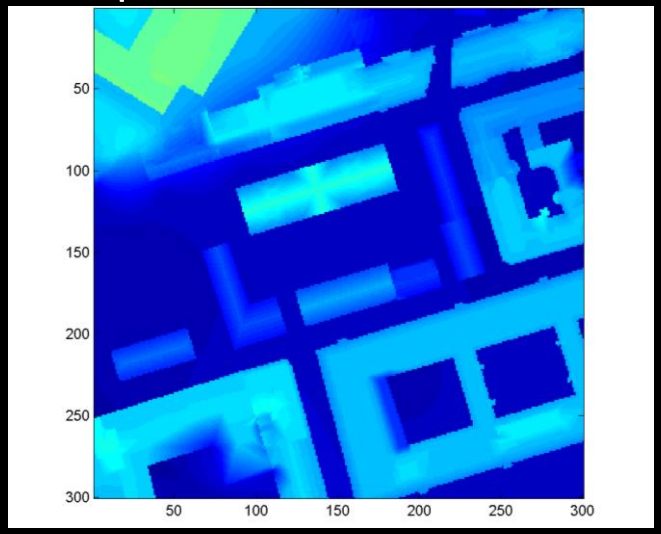
A raster can contain more than one band (e.g. red, green, blue). Than, three values/pixel

Vector

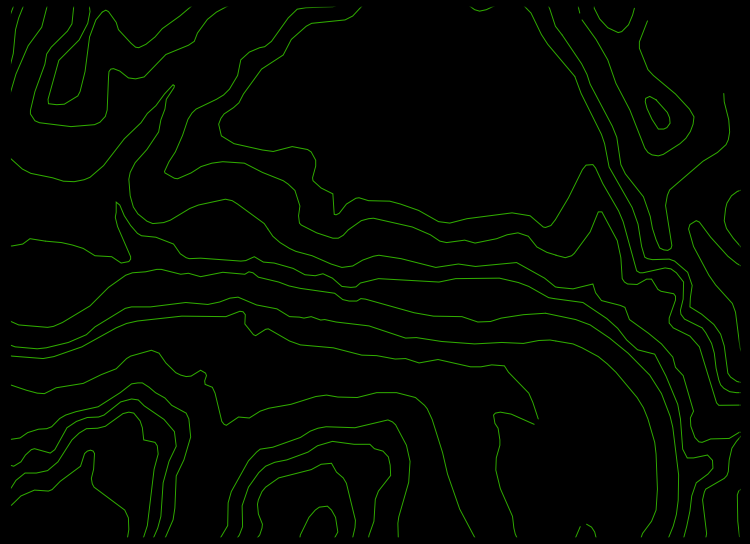
Points

Lines

Polygons (areas)

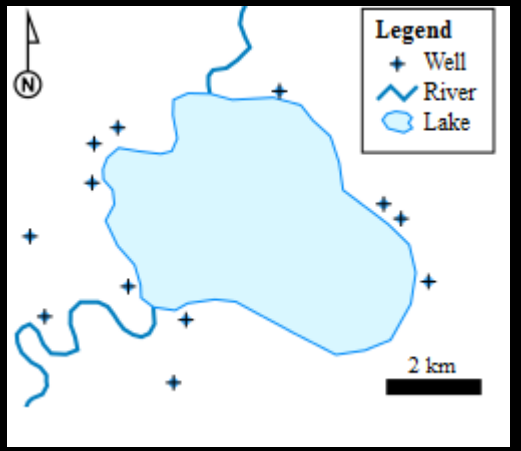


Kronhusbodarna, Göteborg



Contours (isolines)

Data formats



Shapefiles (.shp), vector

Shapefile is a popular geospatial vector data format for GIS software. It is developed and regulated by ESRI as a (mostly) open specification for data interoperability among ESRI and other software products.

Shapefiles spatially describe geometries: points, polylines, and polygons. Each item may also have attributes that describe the items, such as the name or temperature. A Shapefile object could include 3D-info.

Mandatory files :

.shp — shape format; the feature geometry itself
.shx — shape index format; a positional index of the feature geometry to allow fast seeking
.dbf — attribute format; columnar attributes for each shape, in dBase IV format

Example of optional files :

.prj — projection format; the coordinate system and projection information

Data formats

General Markup Language (.gml), vector/raster

GML is the XML grammar defined by OGC to express geographical features. It could hold both vector and raster data. GML contains a rich set of primitives e.g. feature, coordinate reference system, topology etc.

CityGML is a continuation of GML used for 3D objects e.g. Building envelopes

Keyhole Markup Language (.kml)

KML complements GML. KML is a language for the visualization of geographic information tailored for Google Earth. KML instances may be transformed losslessly to GML, however roughly 90% of GML's structures (such as, to name a few, metadata, coordinate reference systems, horizontal and vertical datums, etc.) cannot be transformed to KML. KML-files usually comes as KMZ which is zipped files

XML

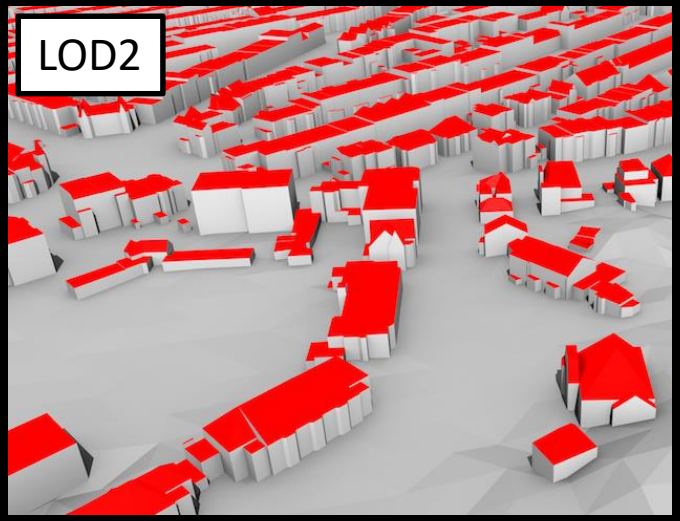
Extensible Markup Language (**XML**) is a set of rules for encoding documents in machine-readable form. The design goals of XML emphasize simplicity, generality, and usability over the Internet.

Data formats

CityGML is a continuation of GML used for *3D objects*, e.g. building envelopes.

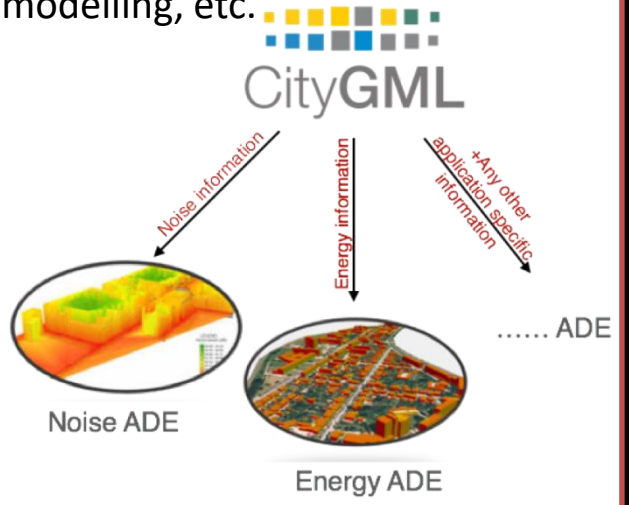
CityGML mainly describes the geometry, attributes and semantics of different kinds of 3D city objects.

Objects can be presented with varying level of detail (LOD), e.g. textures and/or colours.



Application Domain Extensions (ADEs) for CityGML

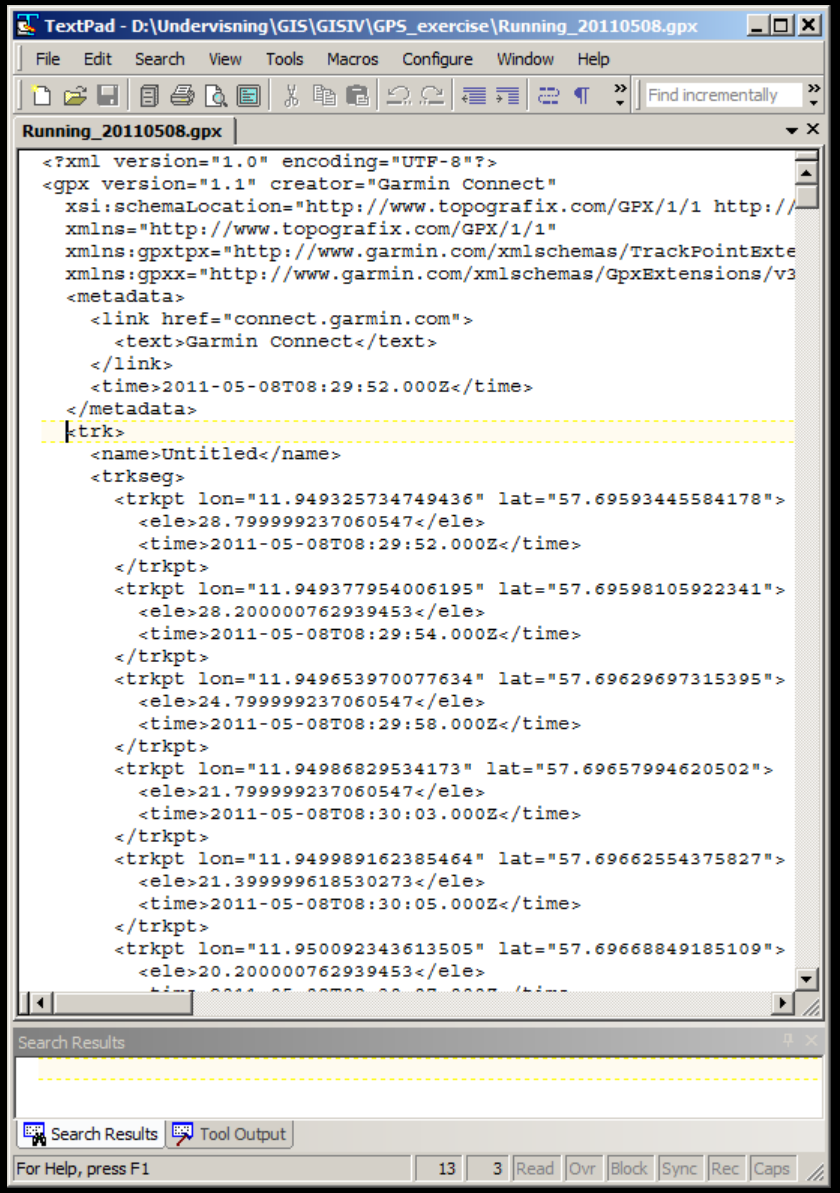
An ADE is defined in an extra XSD (XML Schema Definition) file with its own namespace. ADEs can be defined by information communities which are interested in specific application fields. ADEs are increasingly being used in creating application specific extensions like for energy modelling, modelling topographic data, indoor modelling, noise modelling, etc.



Data formats

GPS data (.gpx), vector

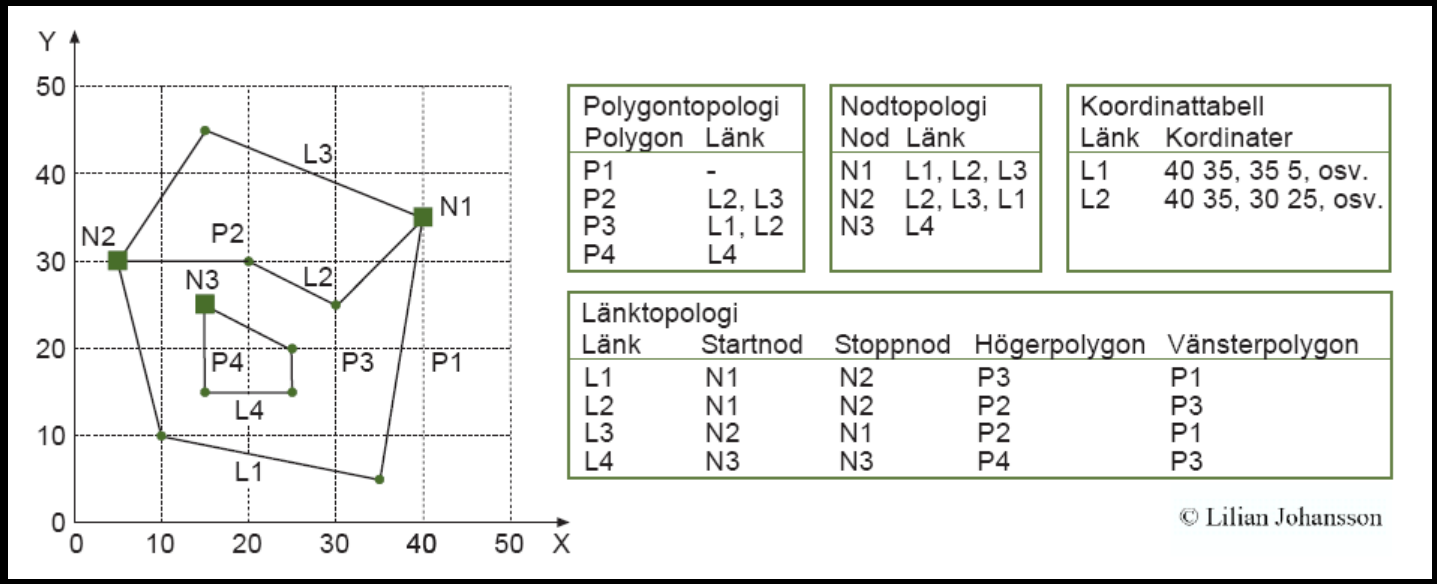
GPS, the Global Positioning System, is a satellite-based system that allows anyone with a GPS receiver to find their exact position anywhere in the world. It is used as an aid in navigation, for example in airplanes, in boats and by hikers. The GPS receiver uses the signals from the satellites to calculate its latitude, longitude and (sometimes) elevation. Most receivers also have the capability to store locations (known as **waypoints**), sequences of locations that make up a planned **route** and a tracklog or **track** of the receivers movement over time. Waypoints, routes and tracks are the three basic feature types in GPS data. This format is also **xml configured**.



Data formats

The GRASS vector data model

GRASS uses a **topological** vector model. This means that areas are not represented as closed polygons, but by one or more boundaries. A boundary between two adjacent areas is **digitized only once**, and it is shared by both areas. Boundaries must be connected and closed without gaps. An area is identified (and labeled) by the centroid of the area. Besides boundaries and centroids, a vector map can also contain points and lines. All these geometry elements can be mixed in one vector and will be represented in different so called ’layers’ inside one GRASS vector map.



Polygontopologi

Polygon	Länk
P1	-
P2	L2, L3
P3	L1, L2
P4	L4

Nodtopologi

Nod	Länk
N1	L1, L2, L3
N2	L2, L3, L1
N3	L4

Koordinattabell

Länk	Kordinater
L1	40 35, 35 5, osv.
L2	40 35, 30 25, osv.

Länktopologi

Länk	Startnod	Stoppnod	Högerpolygon	Vänsterpolygon
L1	N1	N2	P3	P1
L2	N1	N2	P2	P3
L3	N2	N1	P2	P1
L4	N3	N3	P4	P3

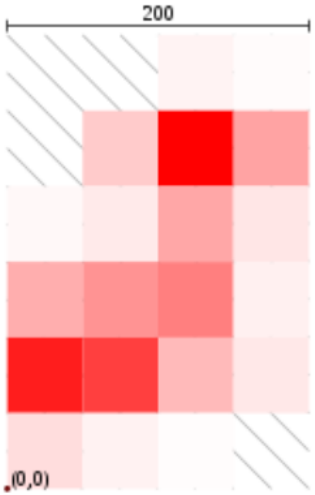
Data formats

ESRI grids, raster

An ESRI grid is a raster GIS file format developed by ESRI, which has two formats:

- A proprietary binary format, also known as an ARC/INFO GRID, ARC GRID and many other variations (.aig)
- A non-proprietary ASCII format, also known as an ARC/INFO ASCII GRID (.asc)

Grid image



Values

	25	75	125	175
275	NA	NA	5	2
225	NA	20	100	36
175	3	8	35	10
125	32	42	50	6
75	88	75	27	9
25	13	5	1	NA

ASCII grid format

```
ncols         4
nrows         6
xllcorner     0.0
yllcorner     0.0
cellsize      50.0
NODATA_value  -9999
-9999 -9999 5 2
-9999 20 100 36
3 8 35 10
32 42 50 6
88 75 27 9
13 5 1 -9999
```

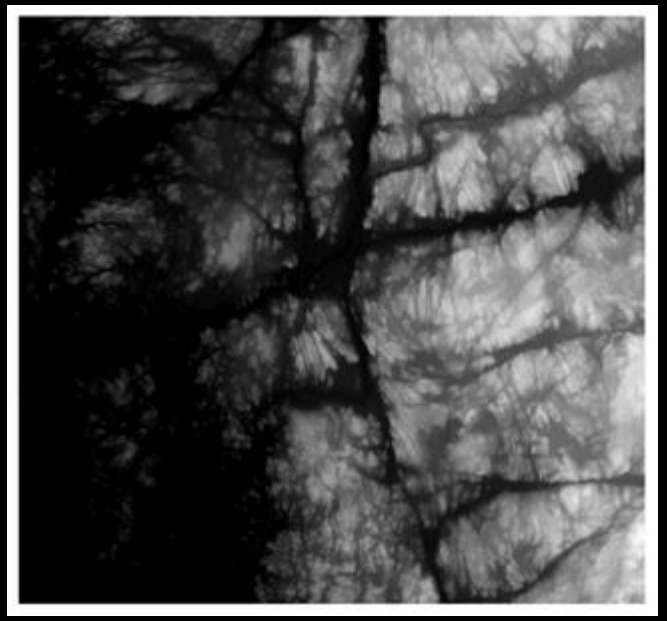
Data formats

GeoTIFF and TIFF (.tif), raster

An image (raster) could be represented by standard image formats e.g. TIFF. A **GeoTIFF** consists of one single file where Coordinate Reference System (CRS) is embedded in the file. GeoTIFF is an open format. If a TIFF file is used, a .tfw-file could be included to hold CRS info. This file is called a **world file**.

JPEG(.jpg), raster

Works in a similar way as a TIFF file.
The world file has the suffix .jpw.



A DEM represented as a TIFF file.

Data formats

GeoPackage

GeoPackage is an open, standards-based, platform-independent, portable, self-describing, **compact format** for transferring geospatial information.

The GeoPackage specification describes a set of conventions for storing the following within an **SQLite database**:

- vector features
- tile matrix sets of imagery and raster maps at various scales
- extensions

These capabilities are built on a common base and an extension Mechanism is described to provide implementors a way to include additional functionality in their GeoPackages.

Since a GeoPackage is a **database**, it supports direct use, meaning that its data can be accessed and updated in a "native" storage format without intermediate format translations. GeoPackages are interoperable across all enterprise and personal computing environments, and are particularly useful on **mobile devices** like cell phones and tablets in communications environments with **limited connectivity and bandwidth**.

Data formats (Web)

Web-Deliverable Data

Since the use of spatial data for web applications has increased dramatically over the last decade, OGC has included a number of standards web based **data formats designed for the web**.

This category of data is often referred to as **W*S**.

Many of desktop OSGIS includes support for al least some of these formats. This allows us to include data from across the Internet in our mapping projects.

Web Mapping Service (WMS)

WMS provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases. A WMS request defines the geographic layer(s) and area of interest to be processed. The response to the request is ***one or more geo-registered map images*** (returned as JPEG, PNG, etc.) that can be displayed in a browser application.

Data formats (Web)

Web Feature Service (WFS)

WFS provides an interface allowing requests for **geographical features** across the web using platform-independent calls. One can think of geographical features as the "source code" behind a map, whereas the WMS interface or online mapping portals like Google Maps return only an image, which end-users cannot edit or spatially analyze. The XML-based **GML** furnishes the default payload-encoding for transporting the geographic features.

Data manipulation operations include the ability to:

- get or query features based on spatial and non-spatial constraints
- create a new feature instance
- delete a feature instance
- update a feature instance

Web Coverage Service (WCS)

WCS provides an interface allowing requests for geographical coverages across the web using platform-independent calls. The coverages are **objects (or images)** in a geographical area

Data formats (Web)

Web Processing Service (WPS)

The OGC WPS Interface Standard provides rules for standardizing how inputs and outputs (requests and responses) for invoking geospatial processing services, such as polygon overlay, as a web service.

Data formats

Choosing your standard format

Nowadays the compatibility between different formats is relatively straight forward. Still, there could be a number of factors you should consider before choosing what formats to use:

Data management

Aid in creating a logical structure for storing your data. A spatial database is sometimes used

Functionality

Make sure that the format you choose can store and be processed based on yours and others needs

GIS Capabilities

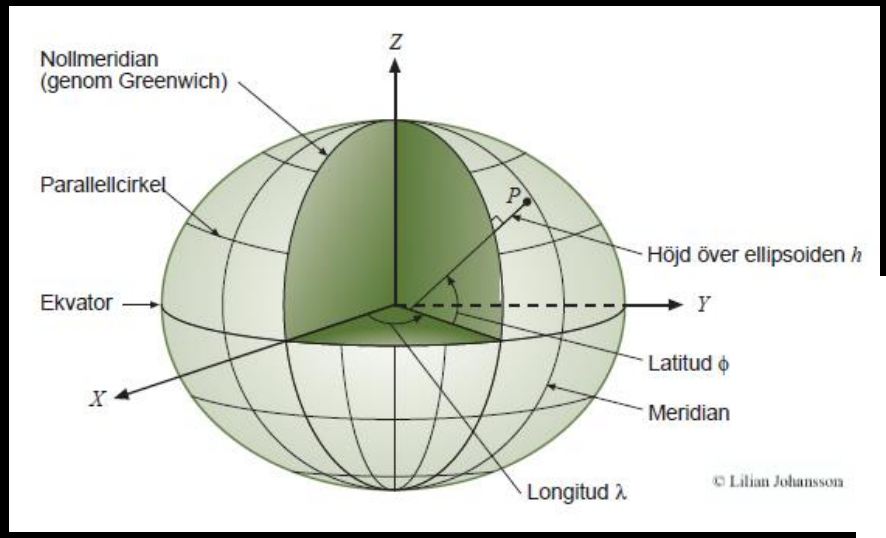
Standard formats is usually in progressing and can be used in new and more advanced GIS analysis

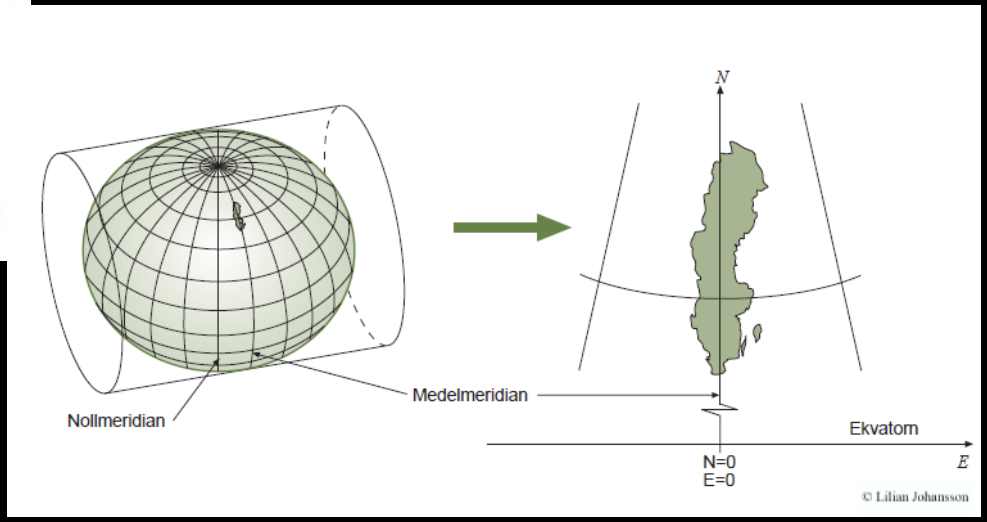
Coordinate Reference Systems (CRS)

Most issues within GIS is related to CRSs

Difference between geographical and projected CRSs?

Always make sure your data and software project are in same CRSs





European Petroleum Survey Group (EPSG)

EPSG through its Geodesy Sub-committee, maintains and publishes a **dataset of parameters for coordinate reference system and coordinate transformation description**. A coordinate system usually has a EPSG ID where all the parameters are represented in order to project a spatial dataset in a GIS software.

WGS84 EPSG ID is 4326 and hold the following information (in .prj format):

GEOGCS["GCS_WGS_1984",DATUM["D_WGS_1984",SPHEROID["WGS_1984",6378137,298.257223563]],PRIMEM["Greenwich",0],UNIT["Degree",0.0174532925199433295]]

Our local CRS, SWEREF99 1200 has the EPSG ID 3007.

Proj.4 and cs2cs conversion

Program *proj.4* is a standard library/program which can be used to *convert geographic longitude and latitude coordinates into cartesian coordinates*, by means of a wide variety of cartographic projection functions. For many of the projection functions the inverse conversion can also be performed (*invproj*).

It is used by many application. E.g. GRASS GIS, PostGIS etc.

The *proj.4* library is limited to converting between geographic and projection coordinates within one datum. The *cs2cs* program operates similarly but allows translation between any pair of definable coordinate systems, including support for datum translation.

datum
a set of parameters that define the position of the origin, the scale, and the orientation of a coordinate system

Tools, data structures and libraries

GDAL/OGR

Using command-line execution is **usually much more effective** and take a fraction of time compared to GUI-based applications.

Runs under the MIT-license (LGPL) which means that it can be in used in proprietary software

GDAL is for raster and **OGR for vector**. Are used “under the hood” of many GIS applications (GRASS, QGIS, Mapguide, FME, ArcGIS). Supports a huge number of formats.

All kind of processing and preprocessing tools are included.

Command-line example:

```
gdalinfo c:\temp\rasterdata.asc
```

Tools, data structures and libraries

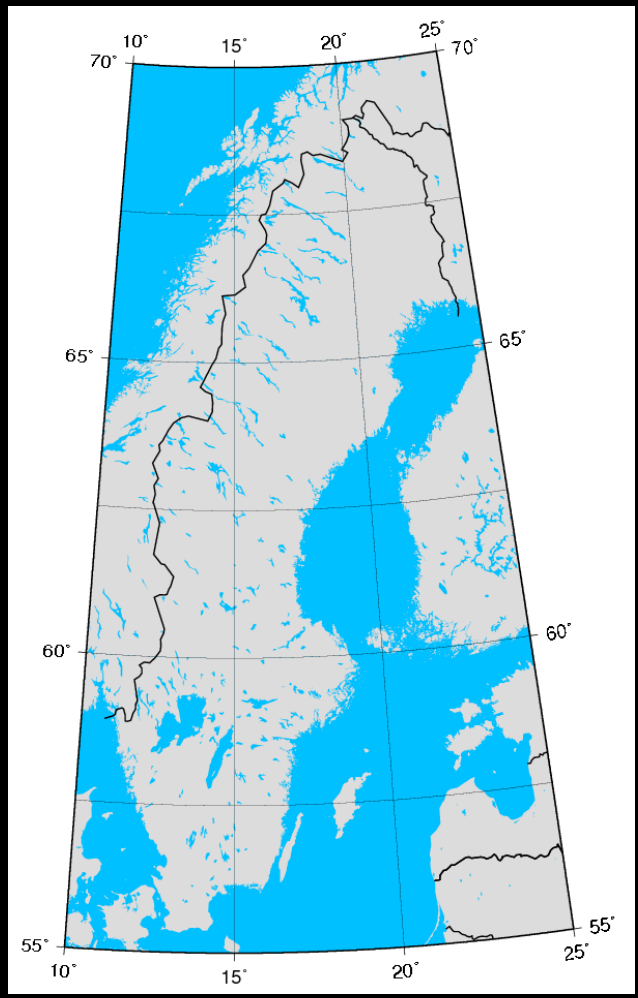
GMT

GMT stands for **the Generic Mapping Tools**. It is also a command-line software which could be used to, among other things, generate overview maps and manipulating geographic and Cartesian data sets. It has a collection of about 65 tools. Some datasets are included which could be used to make maps.

Command-line example:

```
pscoast -JU33N/13 -R10/25/55/70 \
-G220/220/220 -B5g5/5g2.5 -Di -P \
-S0/192/255 -N1/1p/0/0/0 >bild.eps
```

Other, similar Python tool is **cartopy** (before **Basemap**)



Command line execution

Tools that can be executed from a command line interface (CLI) provide a powerful way to manipulate, convert and visualize spatial data. It is also good if you want to process them in batches using a script (upcoming slides). Command line is usually available within the GUI of a software

CLI could be used by systems with insufficient resources to support a graphical user interface (GUI) or to speed up processes since CLI uptake less of a PCs resources.

Command line execution

In Windows

The most common CLI in Windows is ***cmd.exe*** (kommandotolken in Swedish).It can be found under “Tillbehör” (accessories) from the start menu.

Command line execution

IMPORTANT!
The command prompt in windows only recognize the programs that are found in your system path. Same for Mac (IOs) and Linux.

On Windows there you can install Linux like terminal (shell) window e.g. **cygwin** or **msys** that includes the important paths.

OSGeo installations (e.g. QGIS) usually includes its own shell (**OSGeo4W Shell**) where most tools are available.

Command line execution

GRASS
All the components in GRASS can be accessed through a CLI.
Has its own CLI that can be used, ***Text 6.4.1*** (Name depends on version).

Commands can be copied using the GUI.

Example:
g.manual entry=r.info

QGIS
Could be executed from command line using the Python Console included

Scripting in GIS

By combining commands a script (program/plugin) is generated.

Benefits with this approach are:

- A process can be automated and executed very fast.
- Large amounts of data can be processed.
- Repetitive processes can be done very quickly

Different software could have bindings to different computer languages:

GRASS: Shell, Perl, Python etc…
QGIS: Python
ArcGIS: VBScript, Python, JScript, and Perl

Scripting in GIS

By using scripting plugins can be written that could be used in GRASS and QGIS. This is one way that users can contribute to the development of the OS GIS software.

For a closer look on how to create a plugin in in QGIS, see PyQGIS Cookbook online.

GDAL and OGR commands can also be executed using CLI and scripts. This is very powerful when e.g. a whole folder of data layers must be projected from one coordinate system to another.

Open source

The concepts of four freedoms of free software

1. The freedom to run the program, for any purpose.

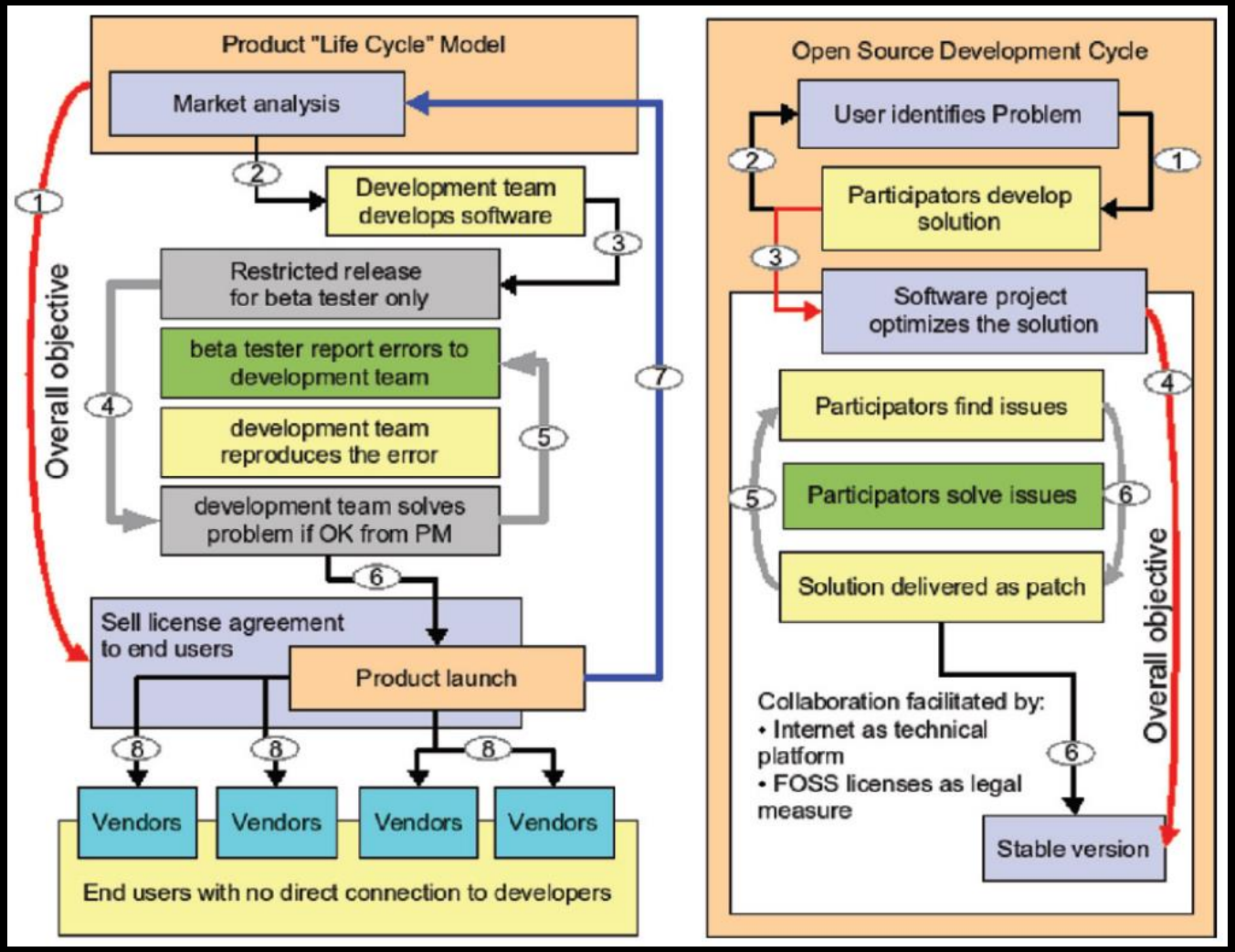
2. The freedom to study how the program works, and adapt it to your needs.

3. The freedom to redistribute copies.

4. The freedom to improve the program, and release your improvements to the public, so that the whole community benefits.

Background

Development cycles for proprietary and OS software



Many OS software starts up and dies because no one is using them

Architecture of most OS Software

A software is usually written in one **computer language** (C++, Java, Python, Ruby, Visual Basic, Perl, …) but usually includes bindings to other languages, especially between C# and Python/Perl/Ruby. The programming language behind an application is important because it affects how the application is distributed, is installed and how easy it is for users to customize.

Most earlier software were based on **command line execution**. **Graphical User Interfaces** (GUI) is now common but still command line can be used.

Usually a **main program core with additional plugins/modules**

Open Source desktop GIS

GRASS GIS

GRASS (Geographic Resources Analysis Support System) is the forefather of OSGIS and is now one of the main projects that OSGeo supports. Developed by the U.S. Army.

Written in C++ but with bindings to e.g. Python. Runs on different platforms.

GRASS is both a GUI-program and a command-line program. The real work is done by a suite of command-line programs, or modules.

PROS:

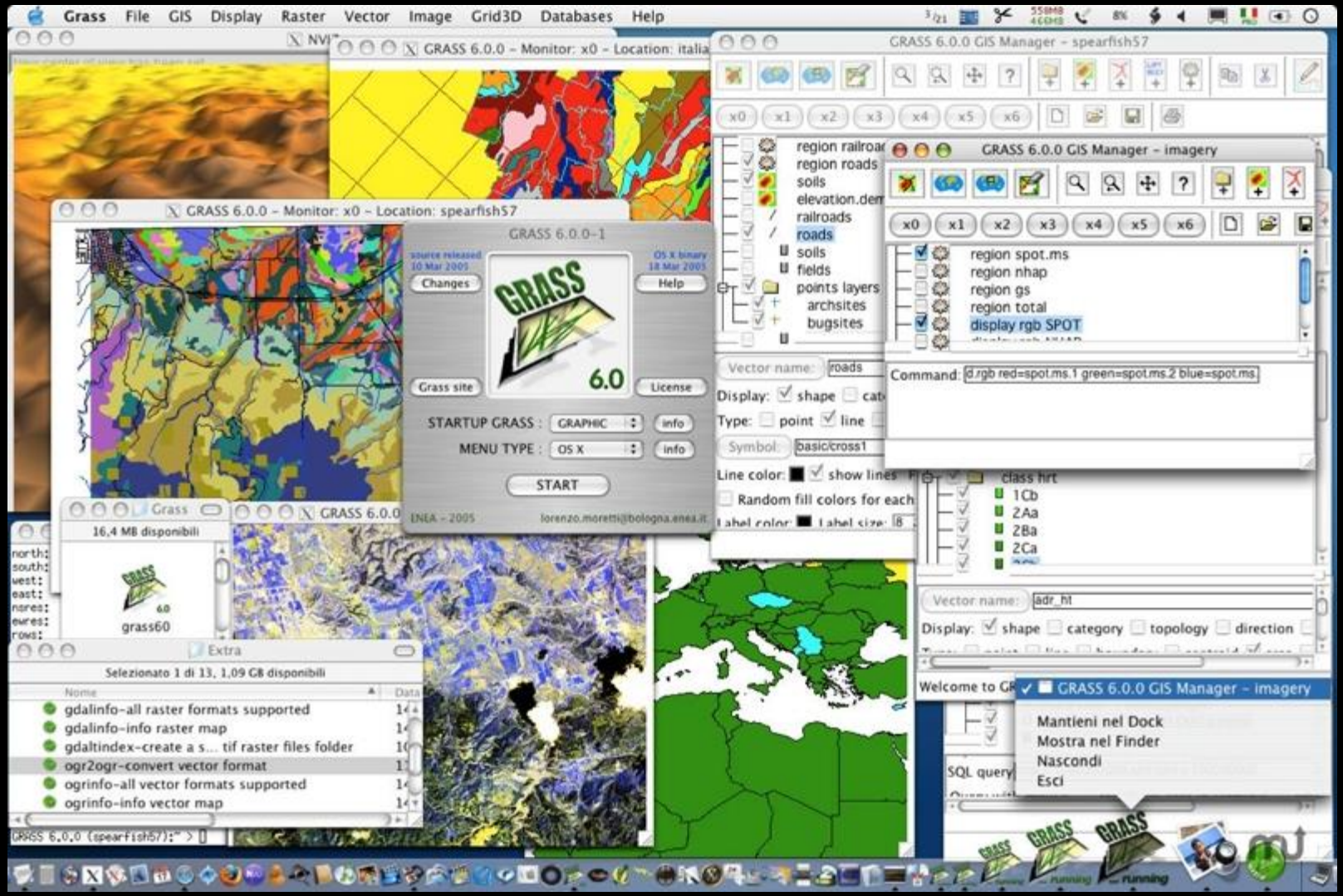
- Stable
- Huge no. of modules
- Good documentation
- 3D visualization
- Raster map algebra and simulation tools
- Can be automated and scripted using common languages

CONS:

- Step learning curve
- Data need to be converted into GRASS format
- “Nonstandard” GUI

Open Source desktop GIS

GRASS GIS



Open Source desktop GIS

gvSIG

A user-friendly OS GIS for both vector and raster. An OSGeo ‘incubation’ project. Developed in Spain.

Written in Java. And runs on different platforms.

gvSIG is being designed to be easily extendable, allowing continuous application enhancement, as well as enabling the development of tailor-made solutions.

PROS:

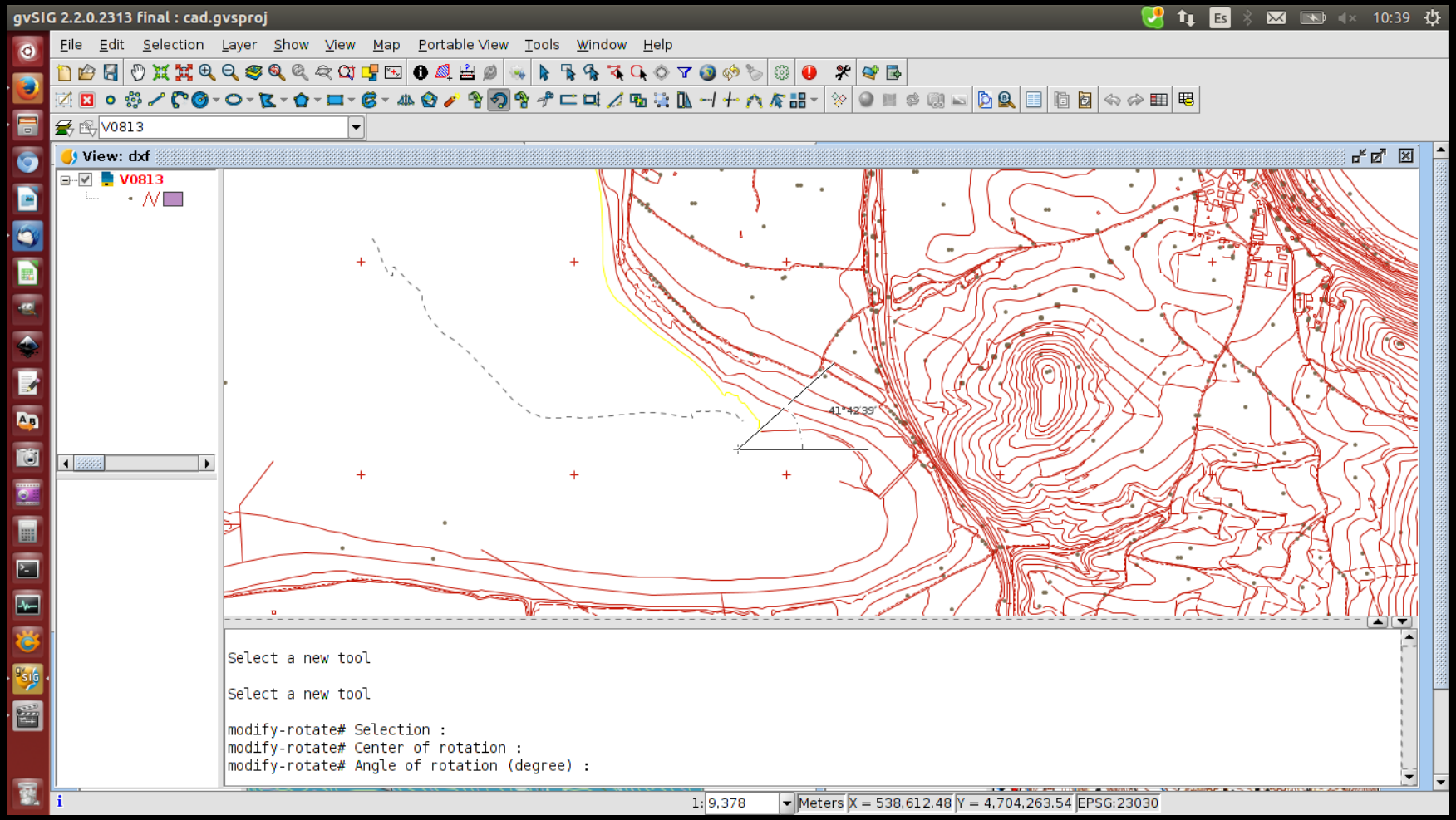
- Good format support
- Extensible through plugins
- Map layouts
- Good geoprocessing tools
- Steep development curve

CONS:

- Still incomplete internationalization (but getting better) (probably not anymore)
- Minor GUI issues on some platforms

Open Source desktop GIS

gvSIG



Open Source desktop GIS

Orfeo ToolBox

An open-source C++ library for remote sensing images processing. Initiated and funded by CNES (French space agency) **Orfeo ToolBox** aims at enabling large images state-of-the-art processing, and is shipped with a set of extensible ready-to-use tools for classical remote sensing tasks, as well as a fully integrated, end-users oriented software called **Monteverdi**.

PROS and CONS:

- Compatible with QGIS

Open Source desktop GIS

SAGA GIS

Stands for **S**ystem for **A**utomated **G**eoscientific **A**nalyses. Originates from the department of Physical Geography in Göttingen University. Written in C++. Based on modules

PROS:

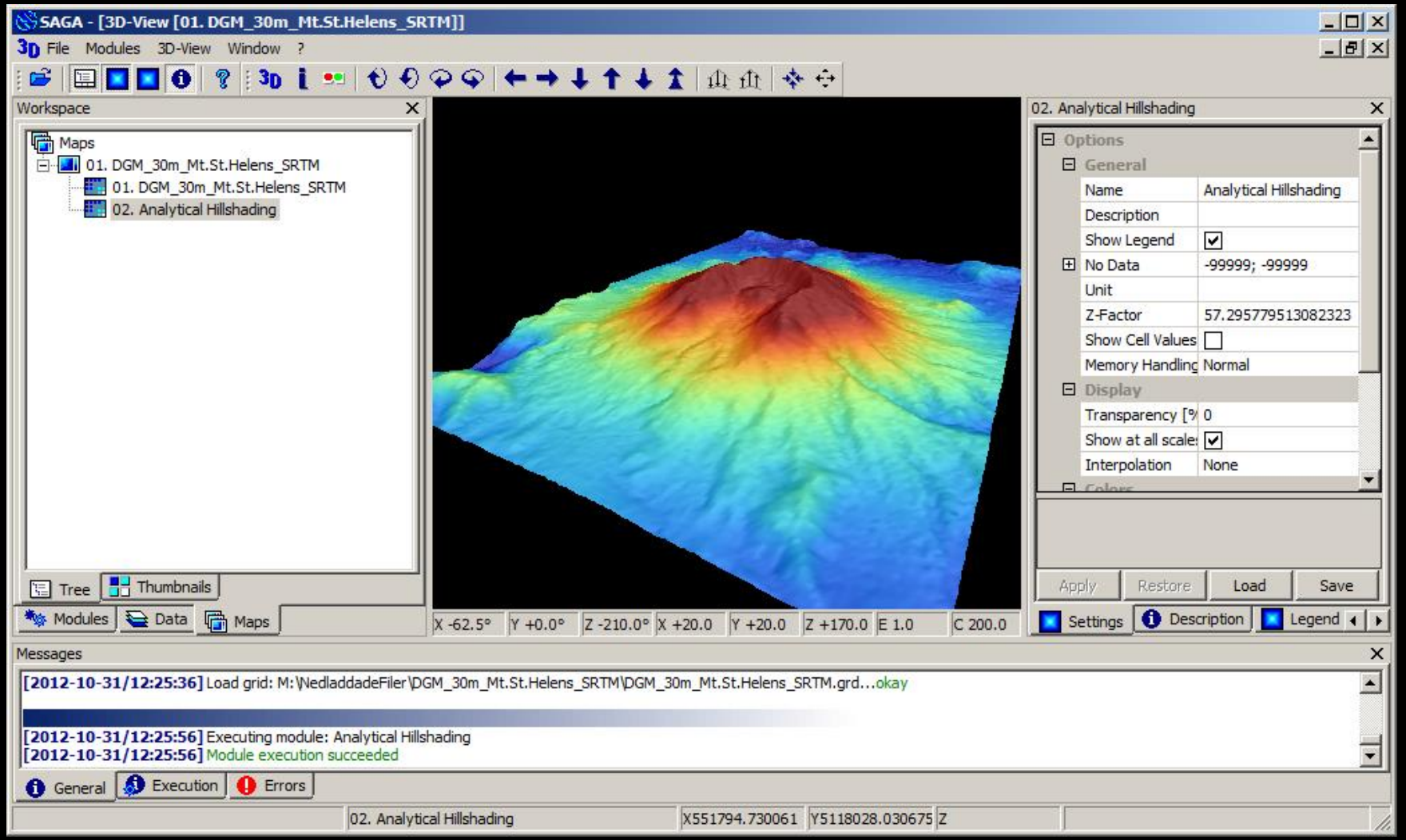
- Specializing on geoscience
- Easy and fast
- 3D compabilities

CONS:

- Don’t know

Open Source desktop GIS

SAGA GIS



Open Source desktop GIS

Marble

Marble is a virtual globe and world atlas — your swiss army knife for maps. Has an globe-style viewer, ossimPlanet (like Google Earth and ArcGlobe).

PROS:

- Available for mobile units (2020?) as well as desktop

CONS:

- Some geodata is not high-resolution

Open Source desktop GIS

Quantum GIS (QGIS)

Founded in 2002. QGIS is an official project of the OSGeo.

QGIS provides a continuously growing number of capabilities provided by core functions and plugins. You can visualize, manage, edit, analyze data, and compose printable maps.

Written in C++ and binds with Python. Parts of the GUI is written in Python. Runs on all platforms. Make use of GDAL/OGR libraries (upcoming slide).

PROS:

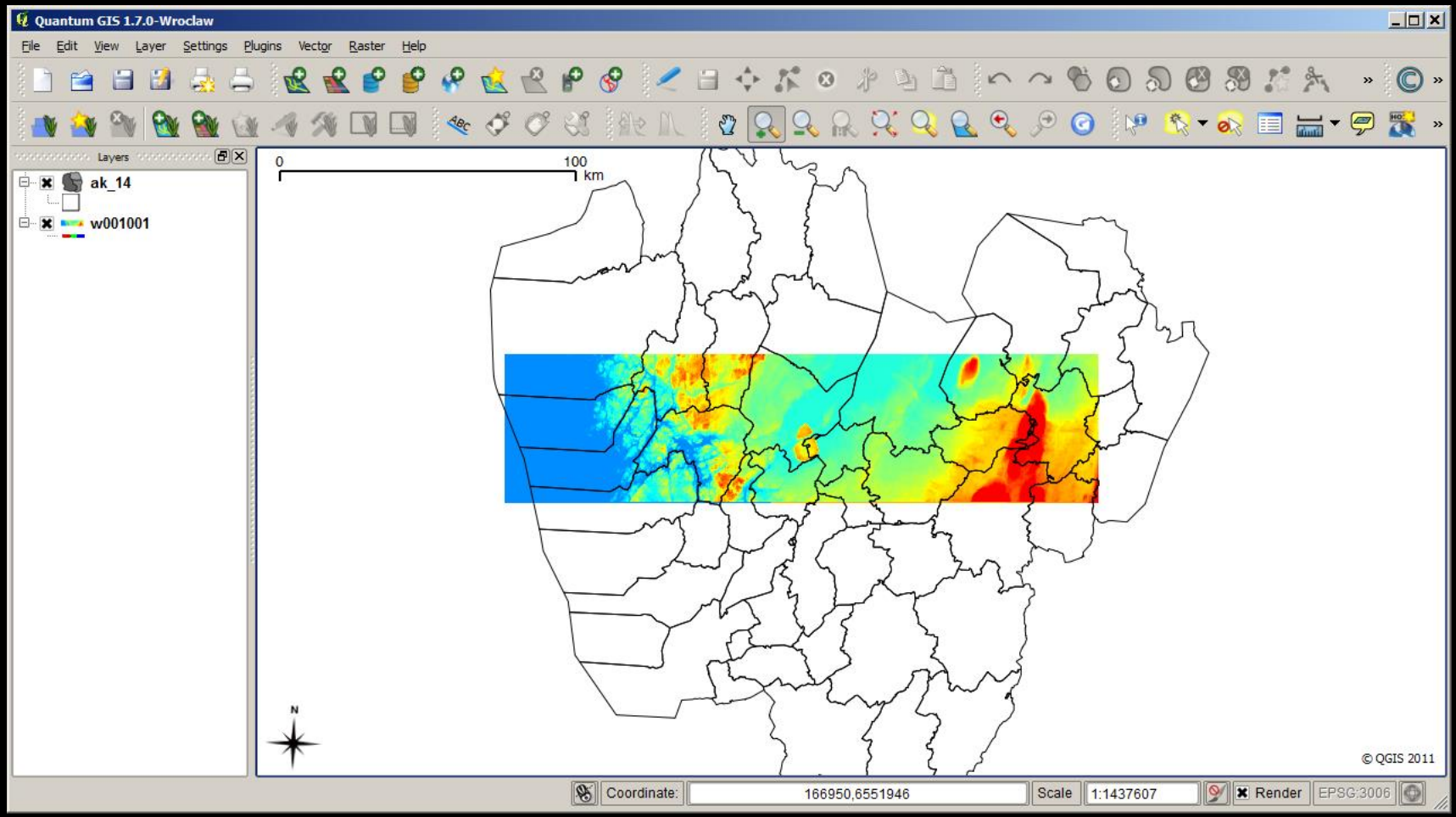
- Supports a wide range of data formats
- Good documentation
- Strong community support (many users)
- Extensible through plugins
- Good integration with GRASS
- Customization using Python

CONS (2012):

- Labeling
- Limited map composition and printing capability
- Still maturing

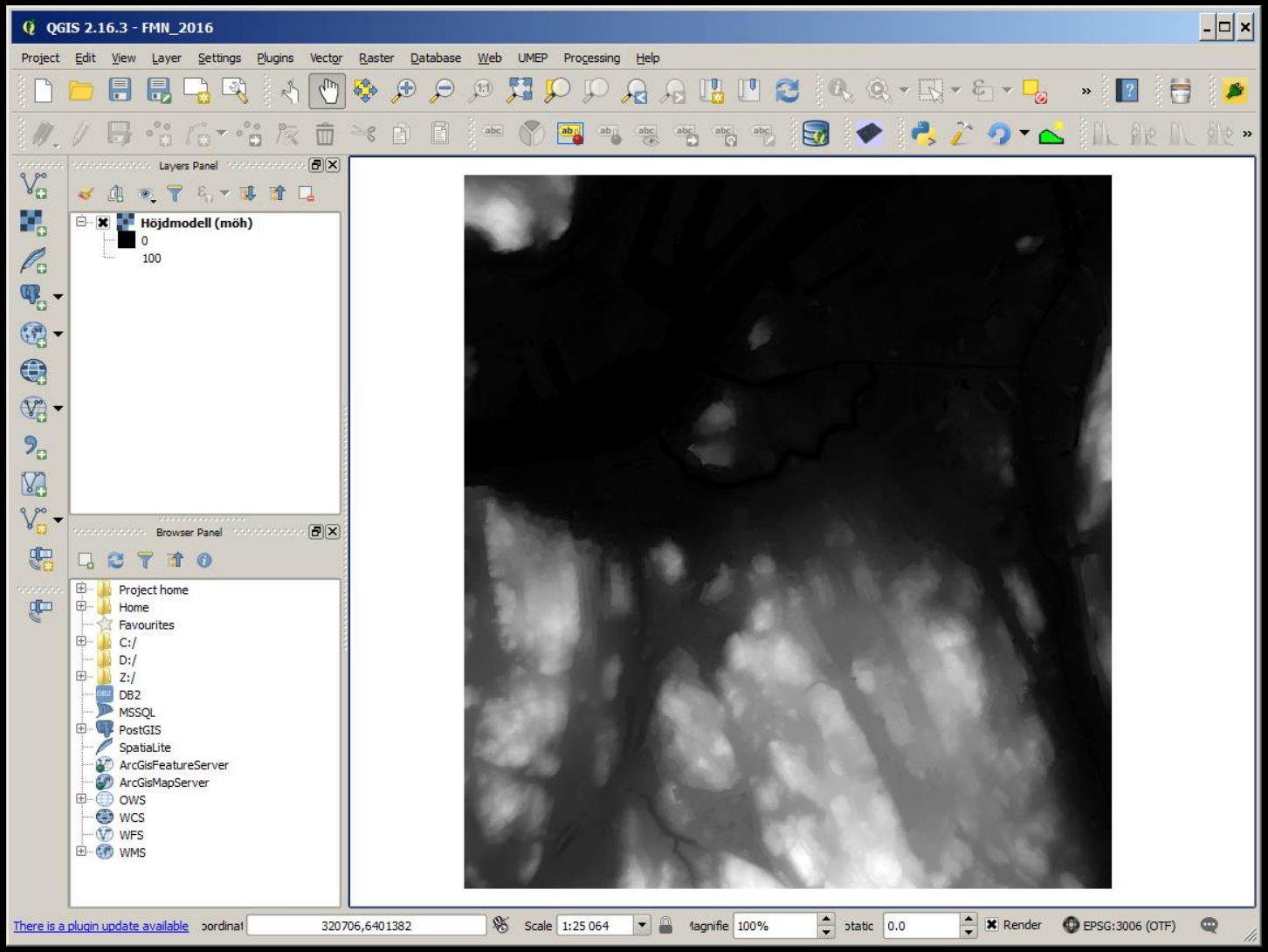
Open Source desktop GIS

Quantum GIS (QGIS)

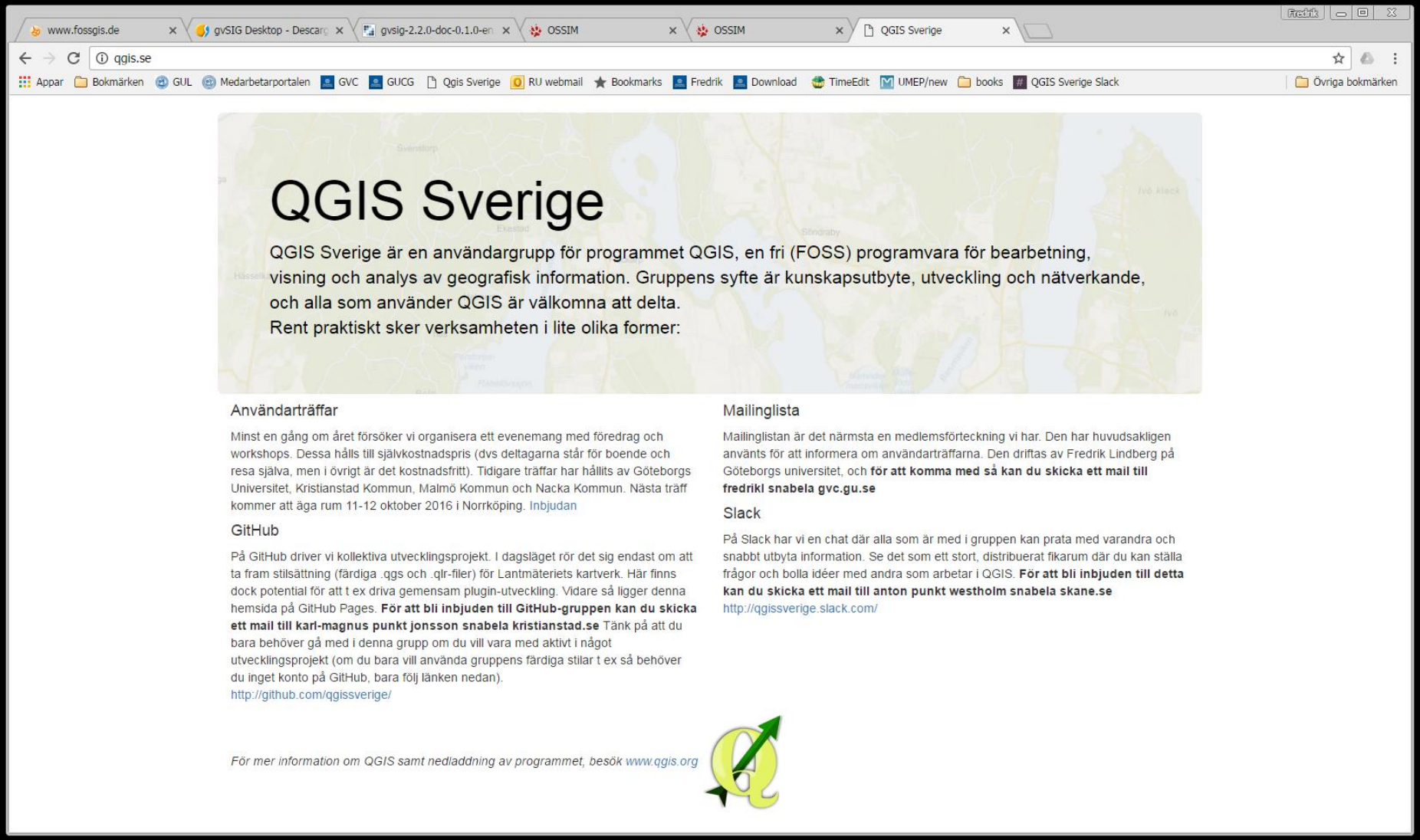


Open Source desktop GIS

Quantum GIS (QGIS)

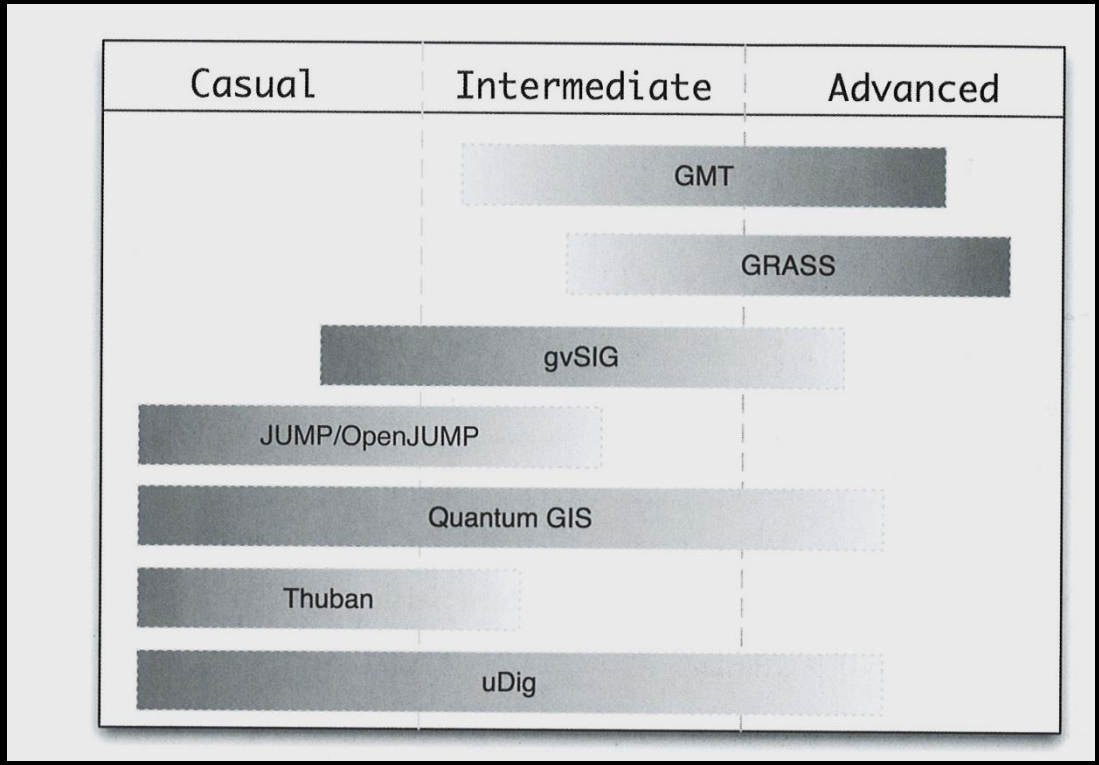


QGIS Sverige



Selecting the right toolkit

Depends on the task and the user



Desktop software development for GIS

Data on search volume over the last 15 year

(trends.google.com)