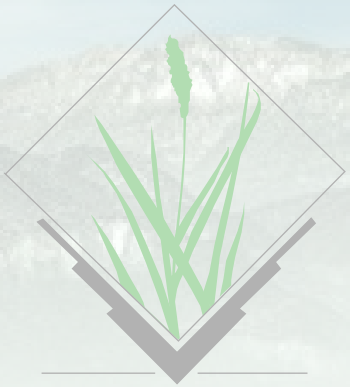
Community based software development: The GRASS GIS project

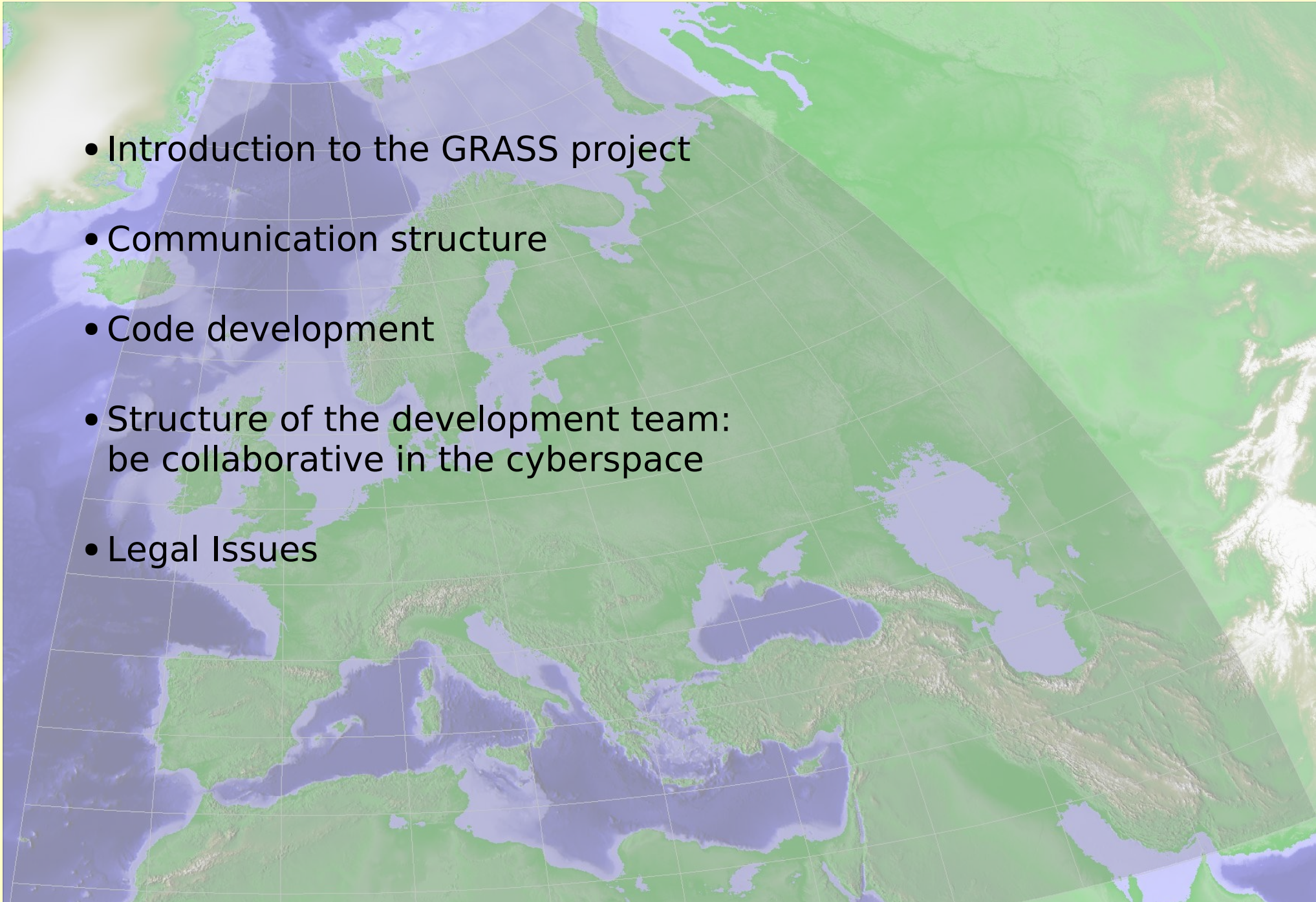
Seminar at
Department of Information and Communication Technology

University of Trento, 24 Apr 2007

M. Neteler
neteler at itc it
http://mpa.itc.it
ITC-irst, Povo (Trento), Italy



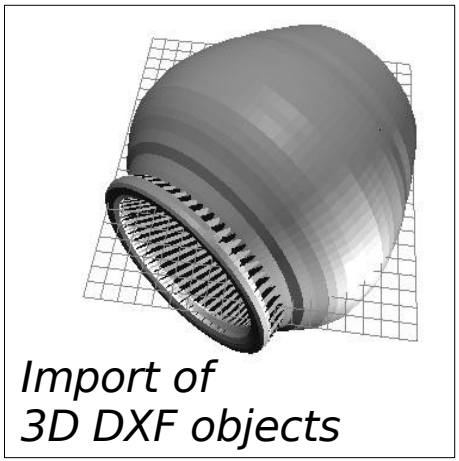
Seminar Outline



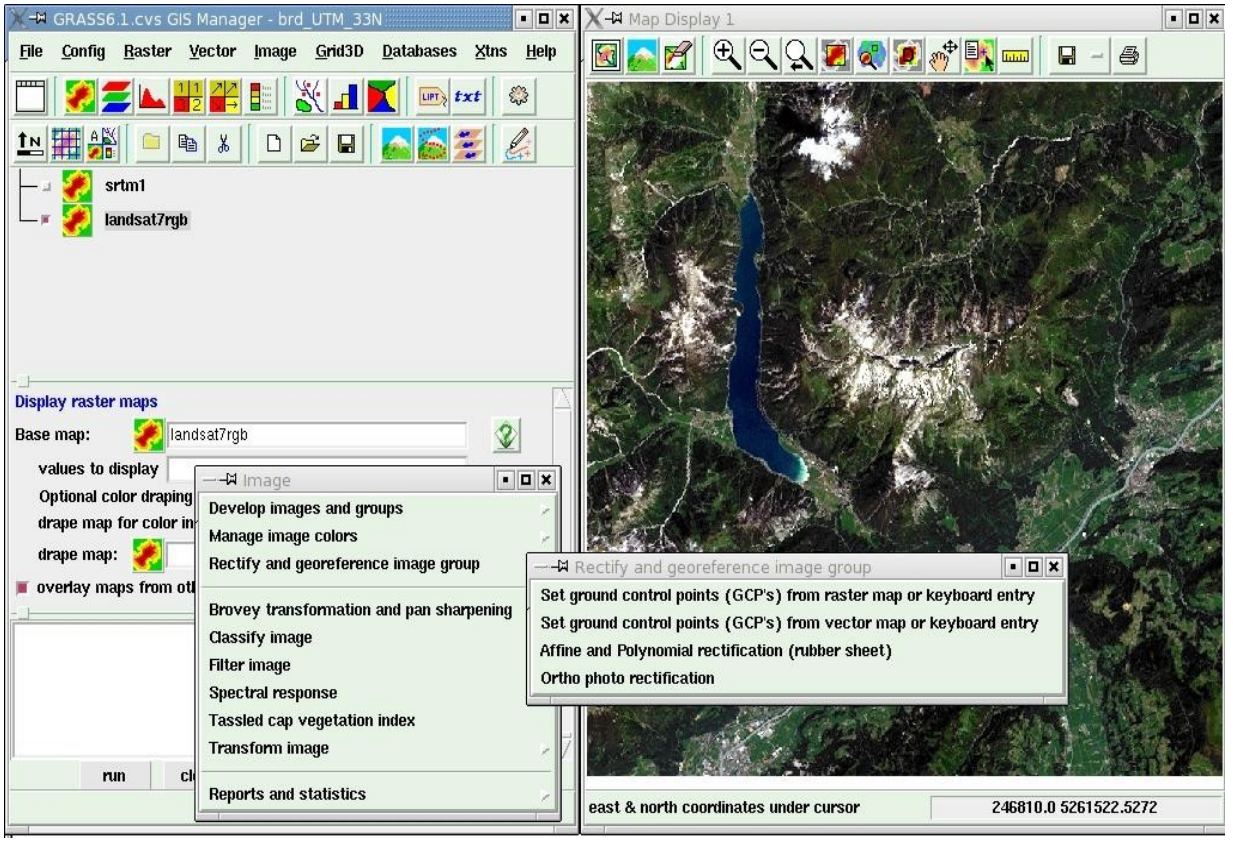
- Introduction to the GRASS project
- Communication structure
- Code development
- Structure of the development team: be collaborative in the cyberspace
- Legal Issues

Objectives of GRASS project

- Continue to develop free software GIS (since 1982)
- Deliver high quality algorithms (often academia based) for
  - spatial data analysis
  - innovative visualization
  - modeling and simulation



Import of 3D DXF objects



Desktop GIS: GRASS GIS

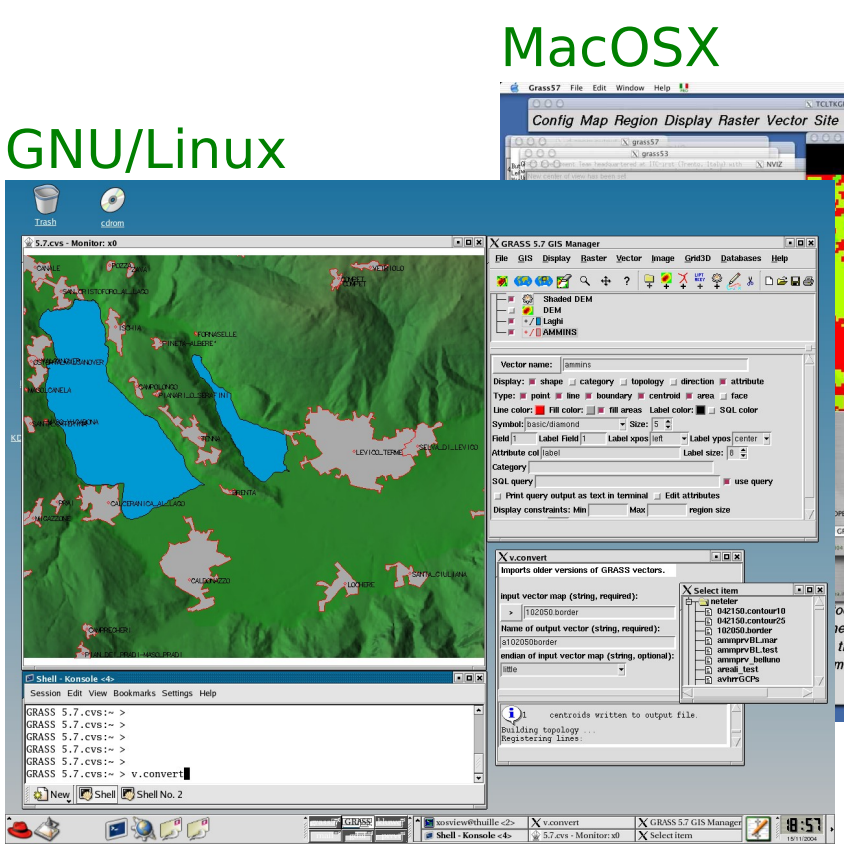
Brief Introduction – Development and System Requirements

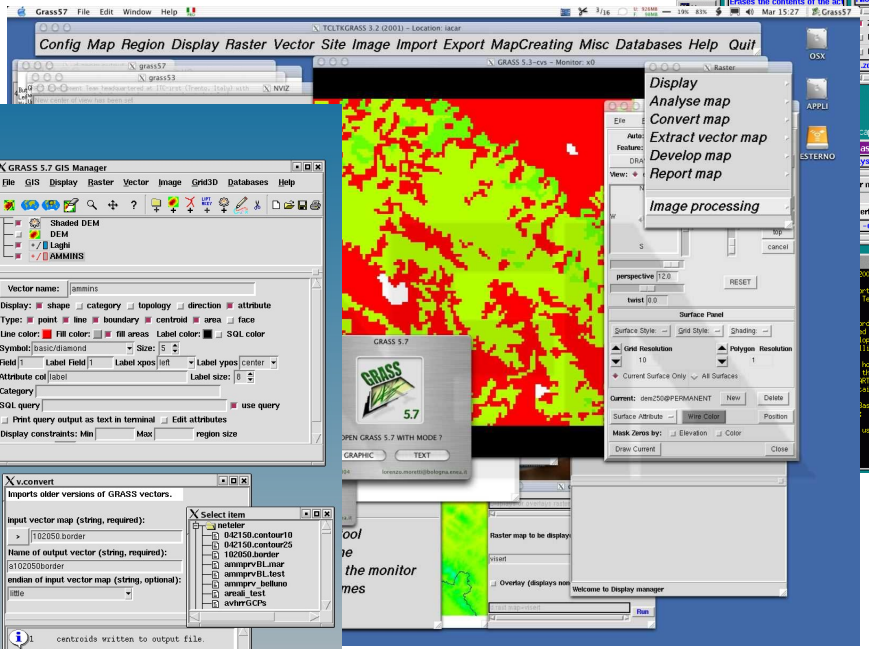
- Developed since 1984, **always Open Source**, since 1999 under GNU General Public License
- Written in C programming language, **portable code** (32/64bit)
- **International development team**, since 2001 coordinated at ITC-irst
- Distributed as source code, precompiled binaries for various platforms, CDROM

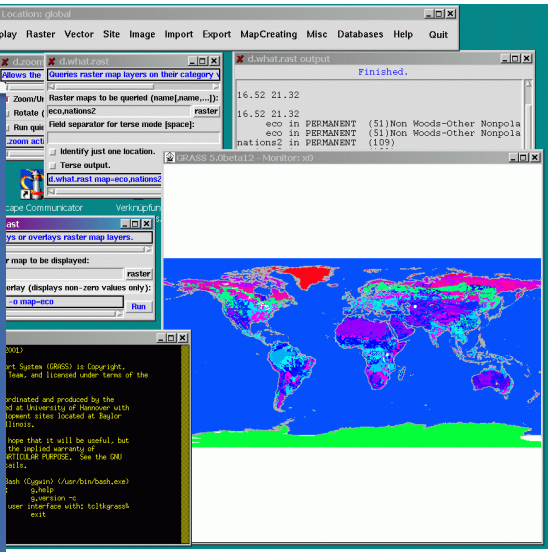
MS-Windows

MacOSX

GNU/Linux







iPAQ

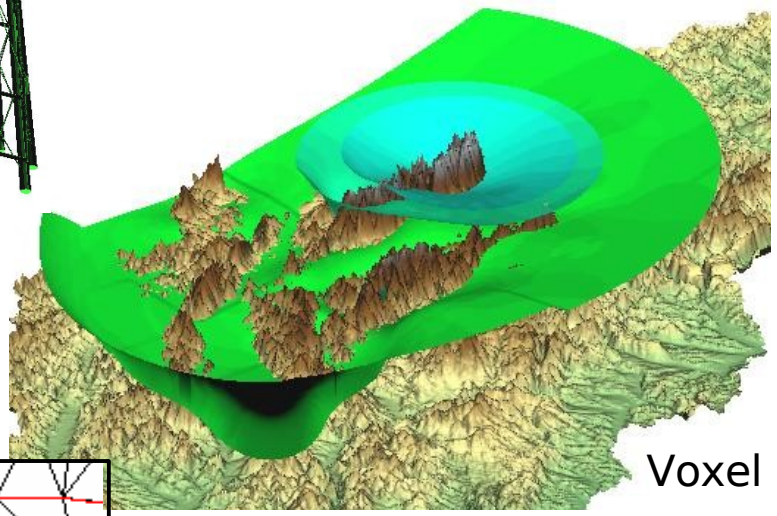
Spatial Data Types

Supported Spatial Data Types

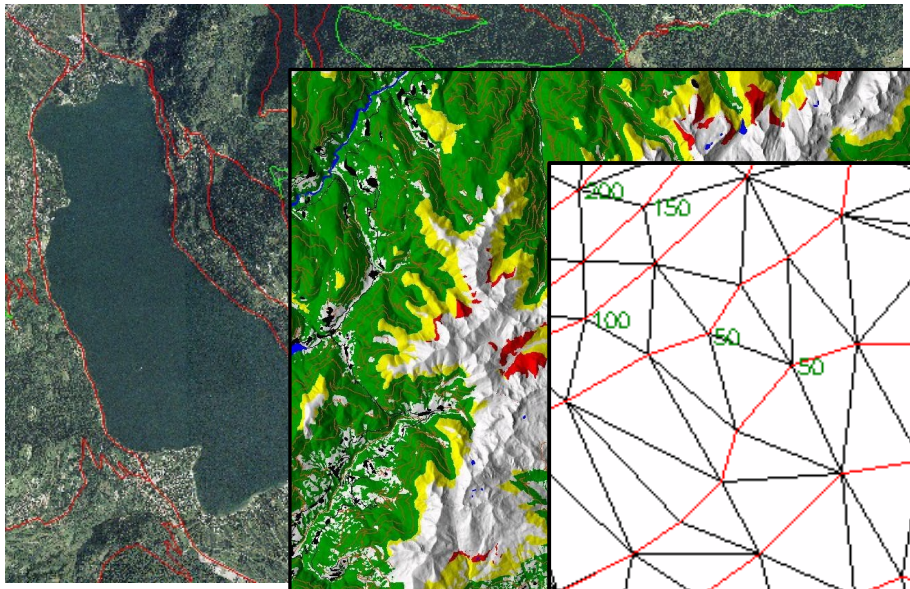
http://grass.itc.it

- 2D Raster data incl. image processing
- 3D Voxel data for volumetric data

- 2D/3D Vector data with topology
- Multidimensional points data



Voxel



Orthophoto

Distances

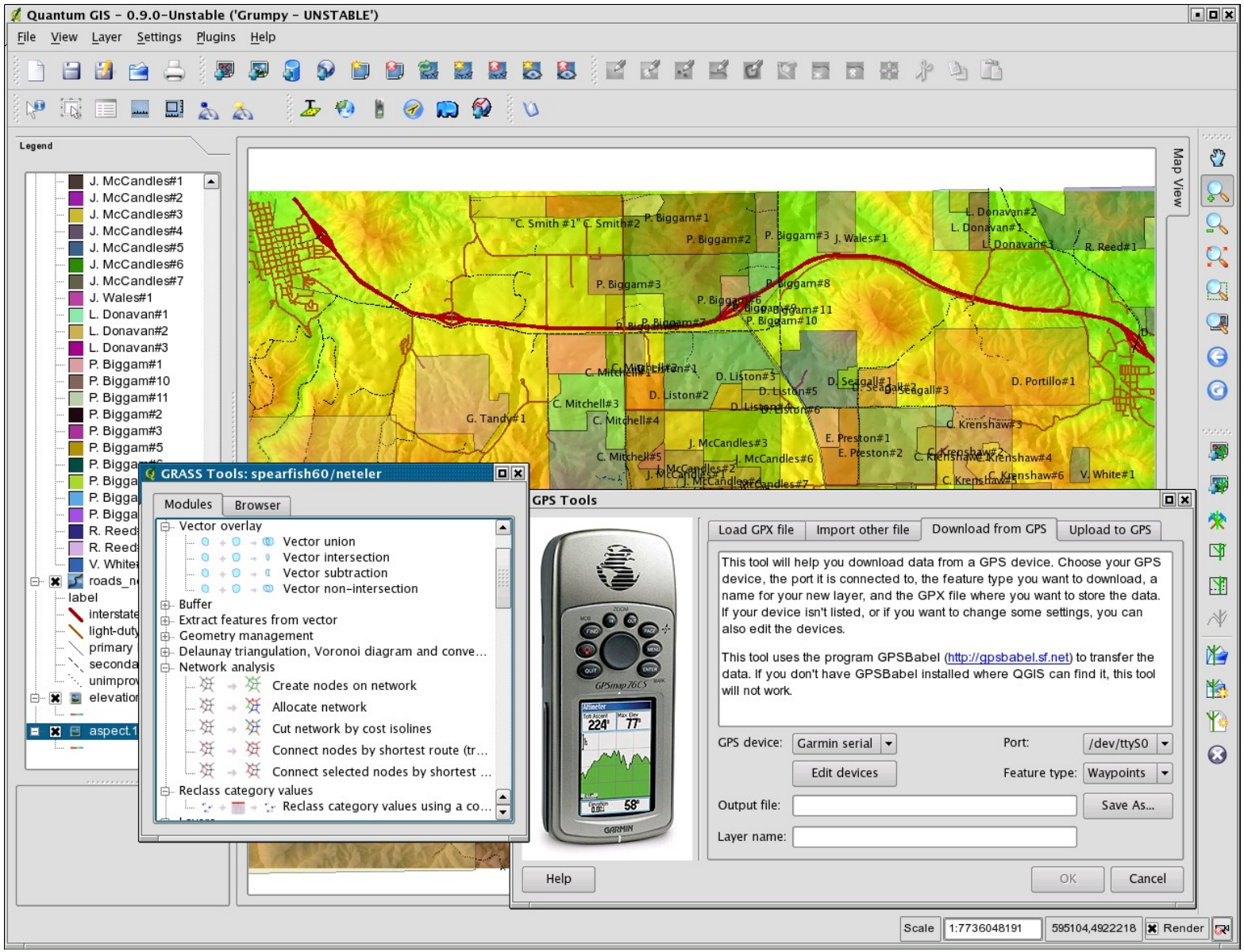
Vector TIN

3D Vector buildings

GRASS GIS

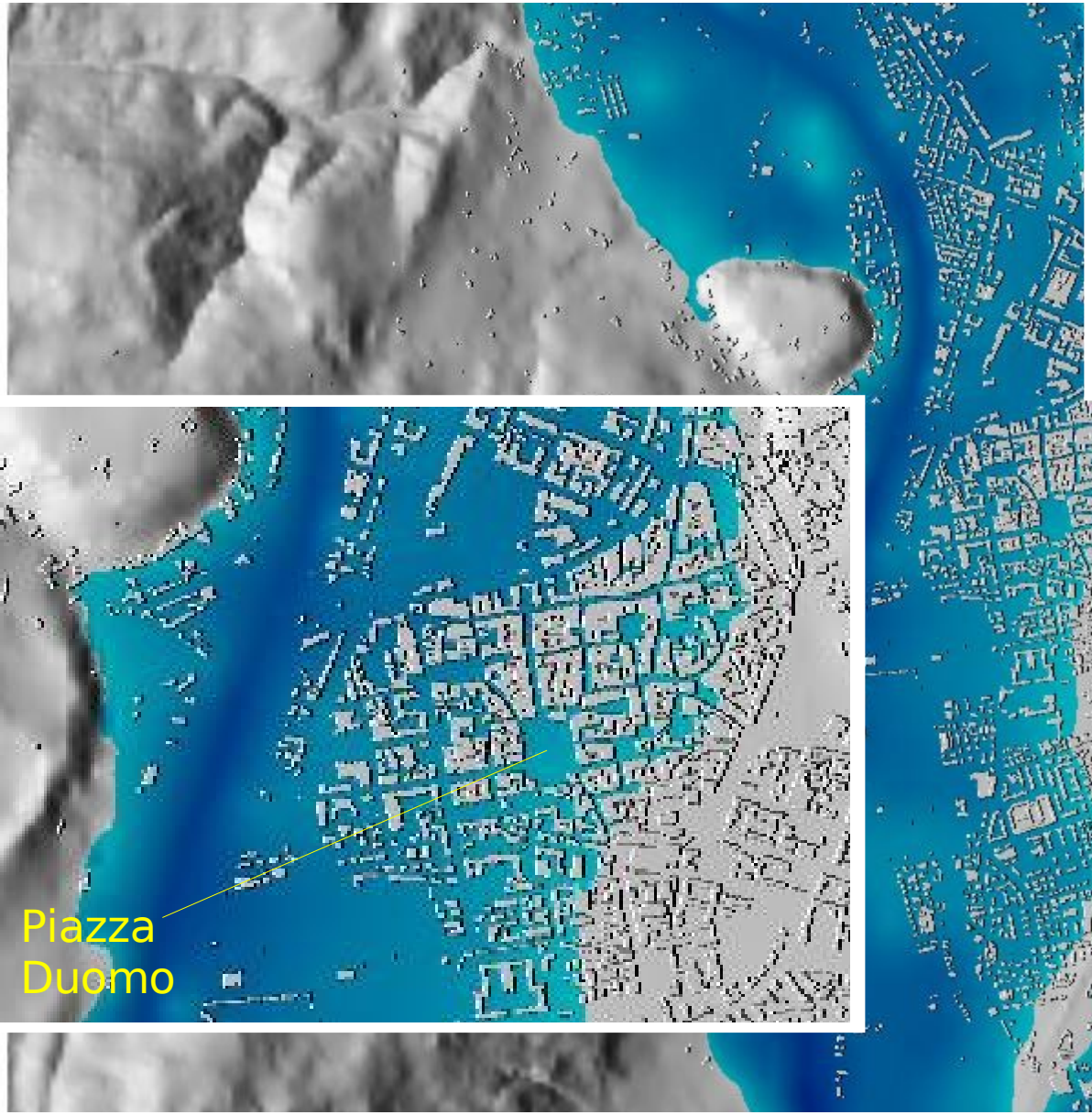
QGIS Geodata viewer with GRASS toolbox, GPS support, Printing Editor ...

http://qgis.org



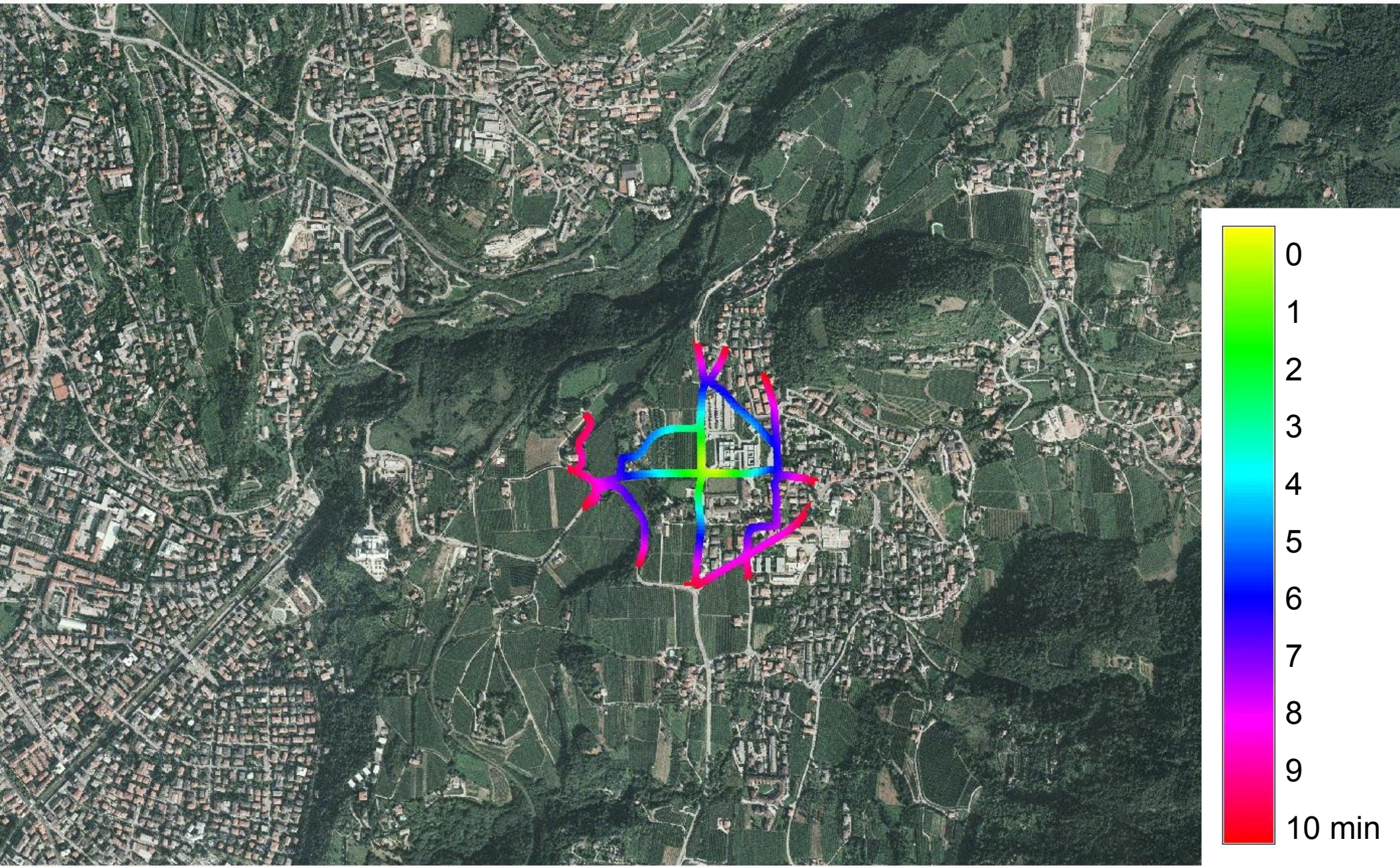
GRASS new features

Flood simulation Trento 1966

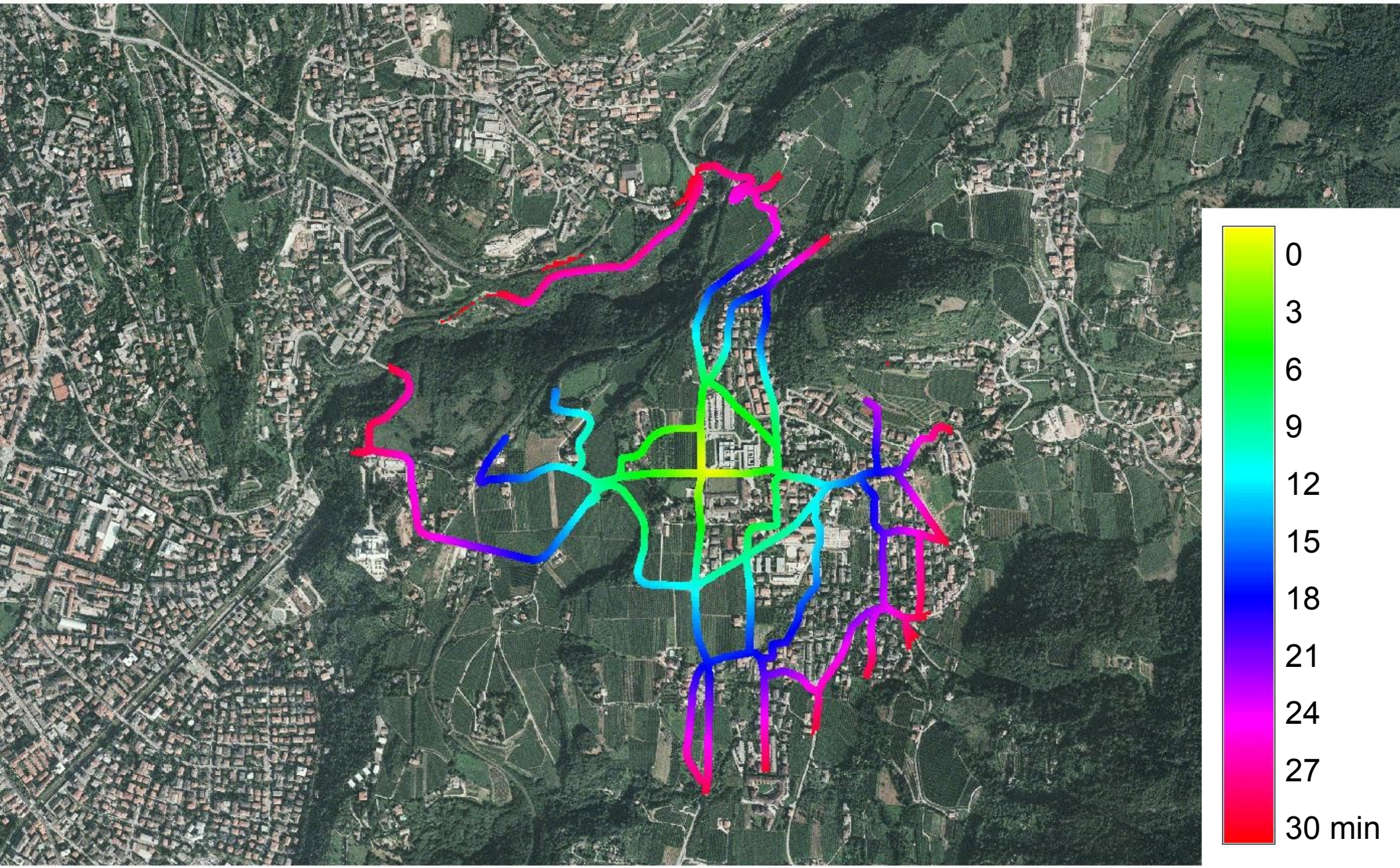


Courtesy: www.questotrentino.it

GRASS: Person walking distance 10 minutes



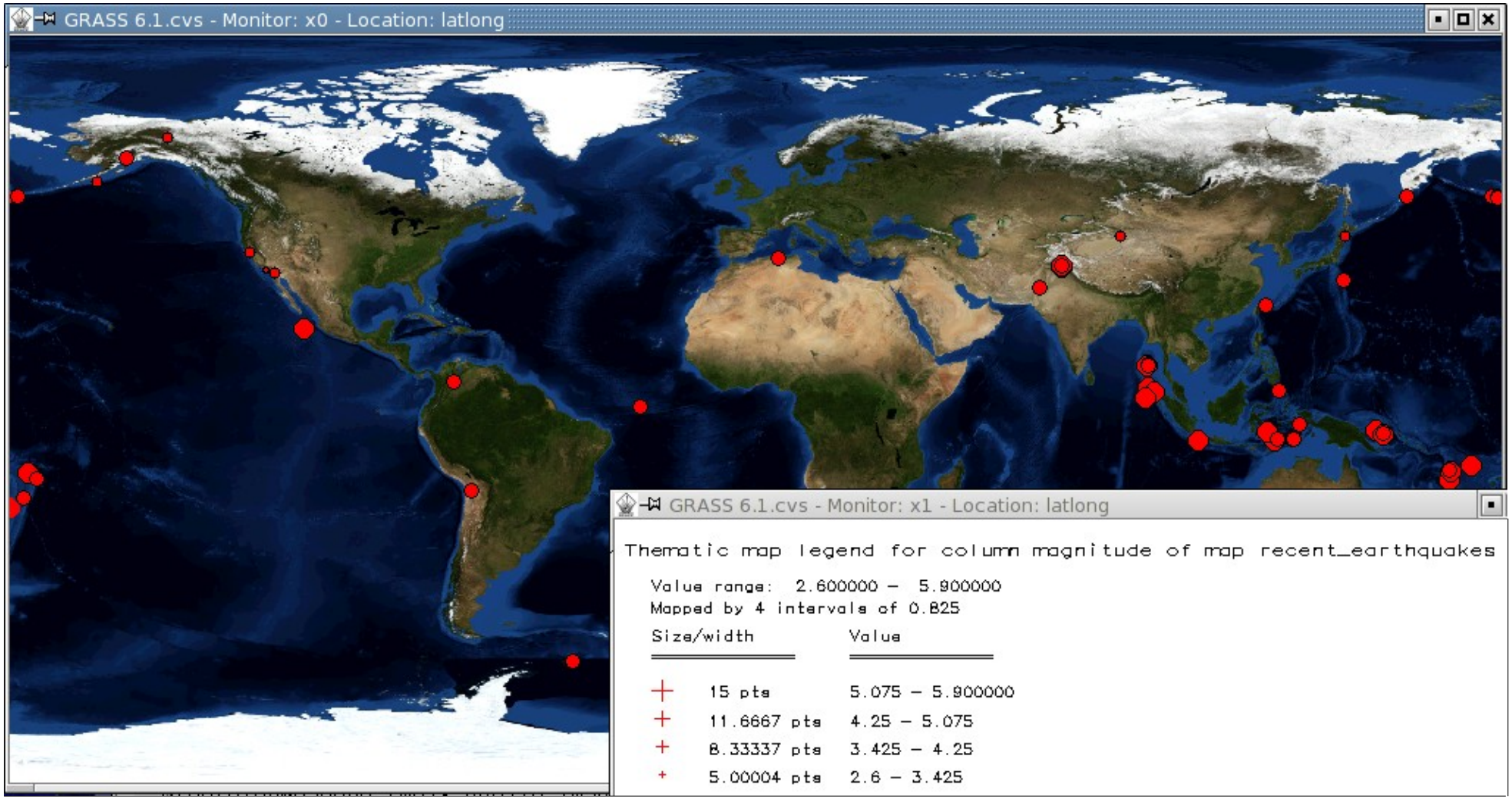
GRASS: Person walking distance 30 minutes



WebGIS: Integration of data sources

GRASS in the Web

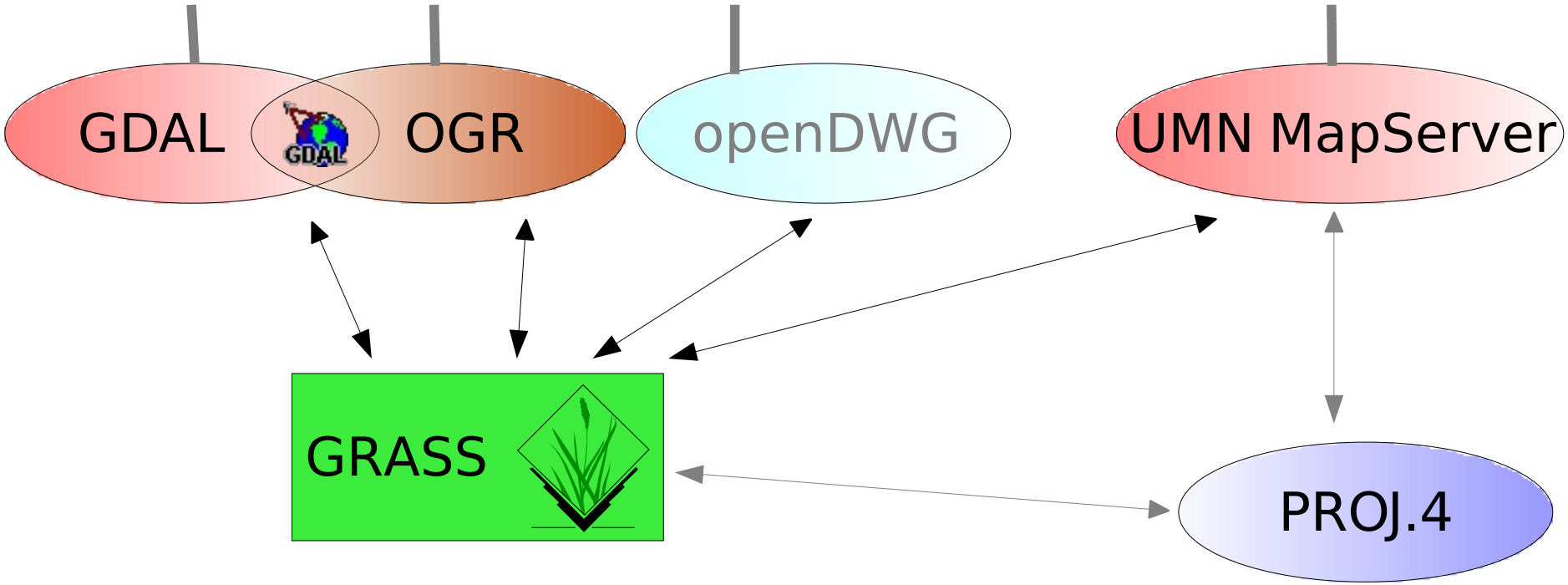
Real-time monitoring of Earthquakes (provided in Web by USGS)
with GRASS/PHP: http://grass.itc.it/spearfish/php_grass_earthquakes.php



GRASS GIS Interoperability

Data models and sources

Raster	Vector	CAD	WebGIS
GeoTIFF	DGN	DXF	Web Map Service (WMS)
Erdas IMG	ESRI-SHAPE	DWG	Web Coverage Service (WCS)
MrSID	GML	...	Web Feature Service (WFS)
ECW	Spatial SQL		Web Map Context Documents (WMC
JPEG2000	...		



GRASS: more than 20 years of free GIS

GRASS 1.0

1984

GRASS 4.1

1993

GRASS 4.2.1/4.3

1998

GRASS 5.0

1999

2000

GRASS 5.1/5.7

2001

GRASS 6.0

2005

U.S. CERL (1984-1995)

1989: civil Internet
1994: first WWW

1997 University of Baylor

1998 University of Hannover

GRASS Development Team
(1997- today)

2001 ITC-irst

PD

GPL'ed

GPL'ed

Manual code management

Automated code management (CVS)

http://grass.itc.it

1990 1992 1994 1997 2004

GRASS Interagency Steering Commitee

Open GRASS Foundation (OGF)

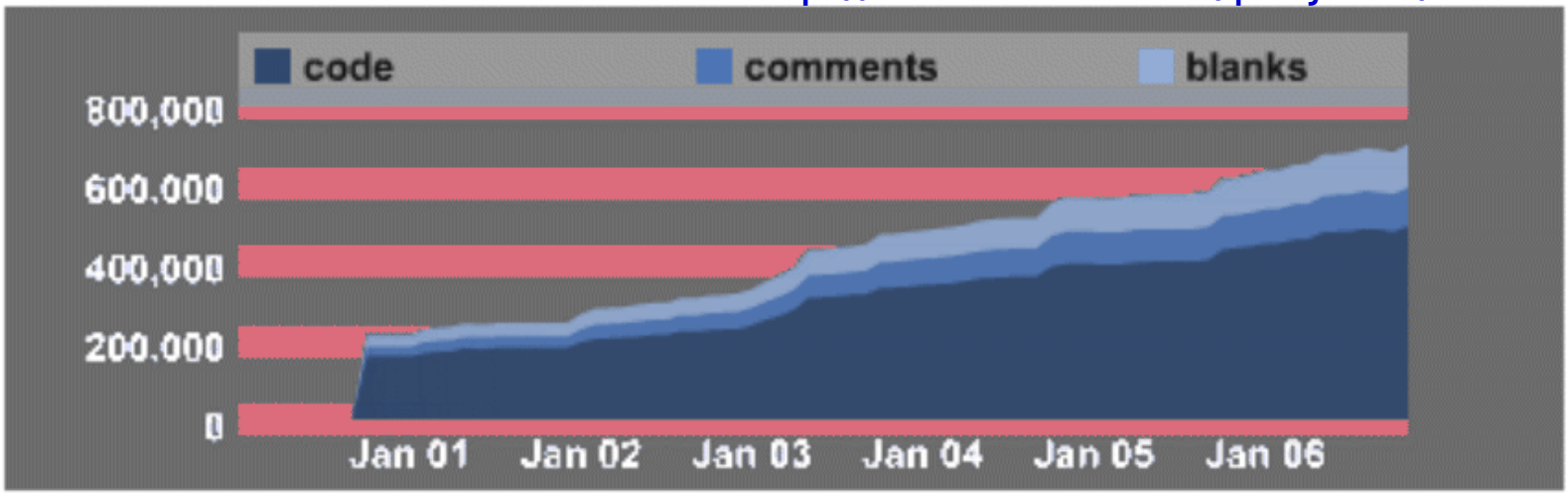
Open GIS Consortium (OGC)

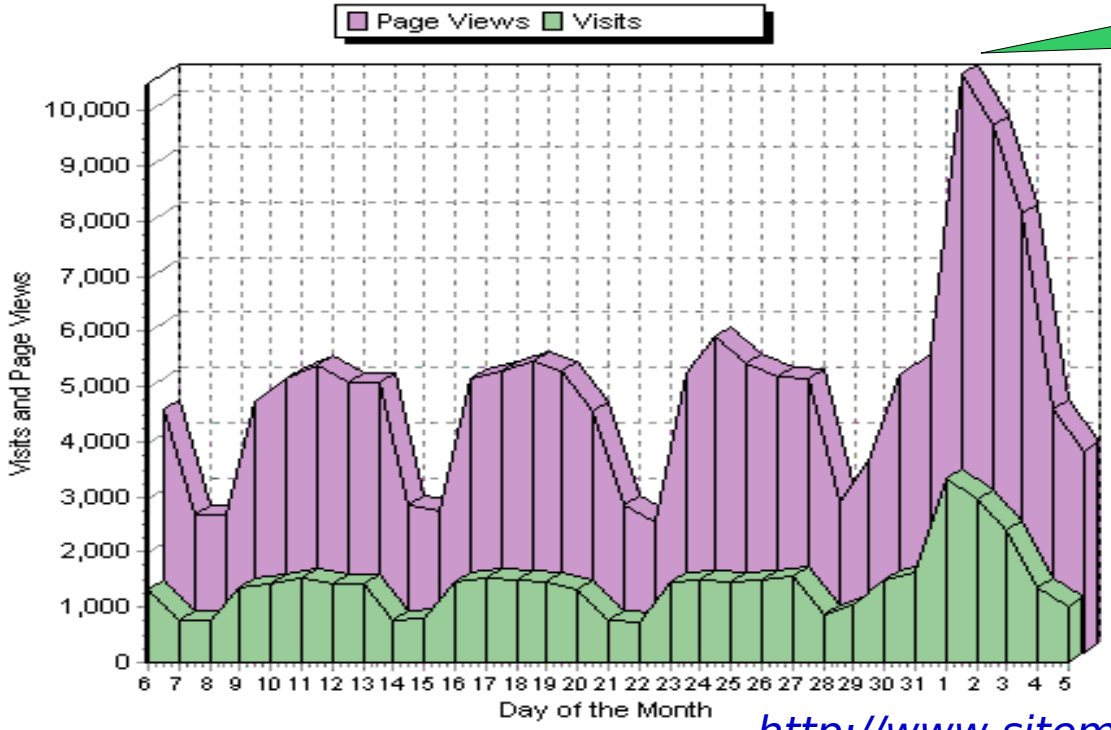
Open Geospatial Consortium (OGC)

GRASS Source Code Statistics

Codebase History **GRASS 6**

http://next.ohloh.net/projects/3666



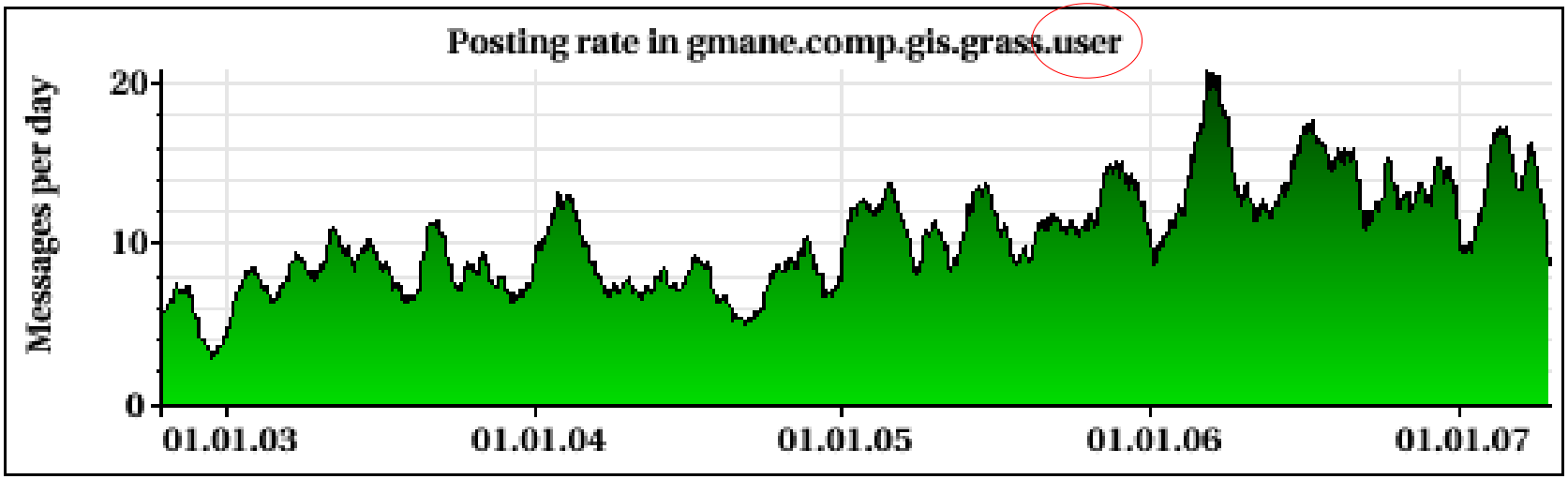


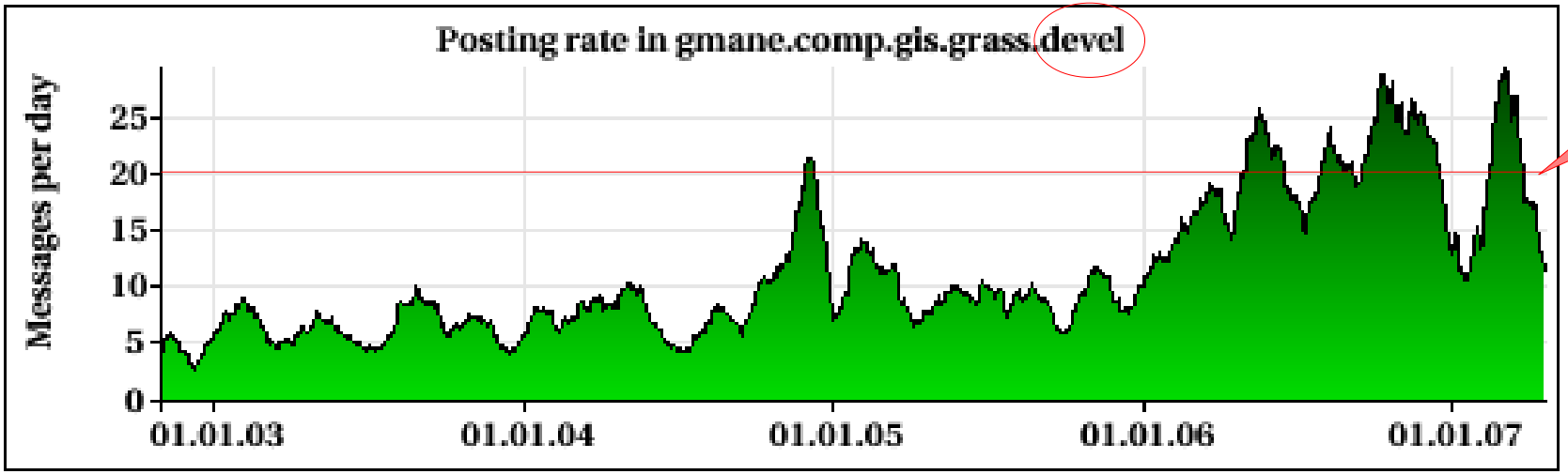
GRASS 6.2.0 release
31 Oct 2006

Visitors at
http://grass.itc.it

http://www.sitemeter.com/?a=stats&s=s24grassgis

GRASS Mailing List Statistics: messages per day





(*http://gmane.org is a mailing list mirror*)

GRASS SLOC Analysis

GRASS 6.3.CVS, 16 Apr 2007

Project Cost

This calculator estimates how much it would cost to hire a team to write this project from scratch.

Include	Markup And Code
Codebase	509,632 LOC
Effort (est.)	137 Person Years
Avg. Salary	$ 55000 /year
	$7,555,370

http://next.ohloh.net/projects/3666

Basic COCOMO model, but slightly different parameters

Or do this analysis yourself - Download and run:
http://www.dwheeler.com/sloccount/

SLOC Totals grouped by language:

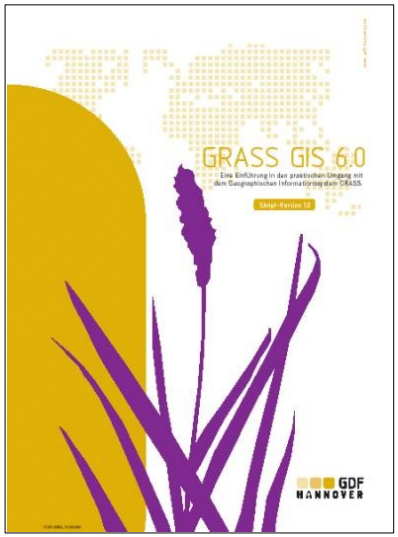
ansic:	473155 (84.30%)
tcl:	44256 (7.88%)
sh:	19821 (3.53%)
python:	10517 (1.87%)
cpp:	10142 (1.81%)
perl:	1608 (0.29%)

...

Total Physical Source Lines of Code (SLOC) = 561,286
Person-Years = 154.05
...
Total Estimated Cost to Develop = $ 20,810,621
(average salary = $56,286/year, overhead = 2.40)

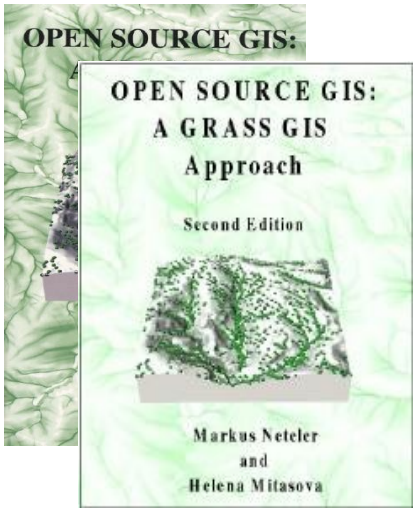
Generated using David A. Wheeler's 'SLOCCount'

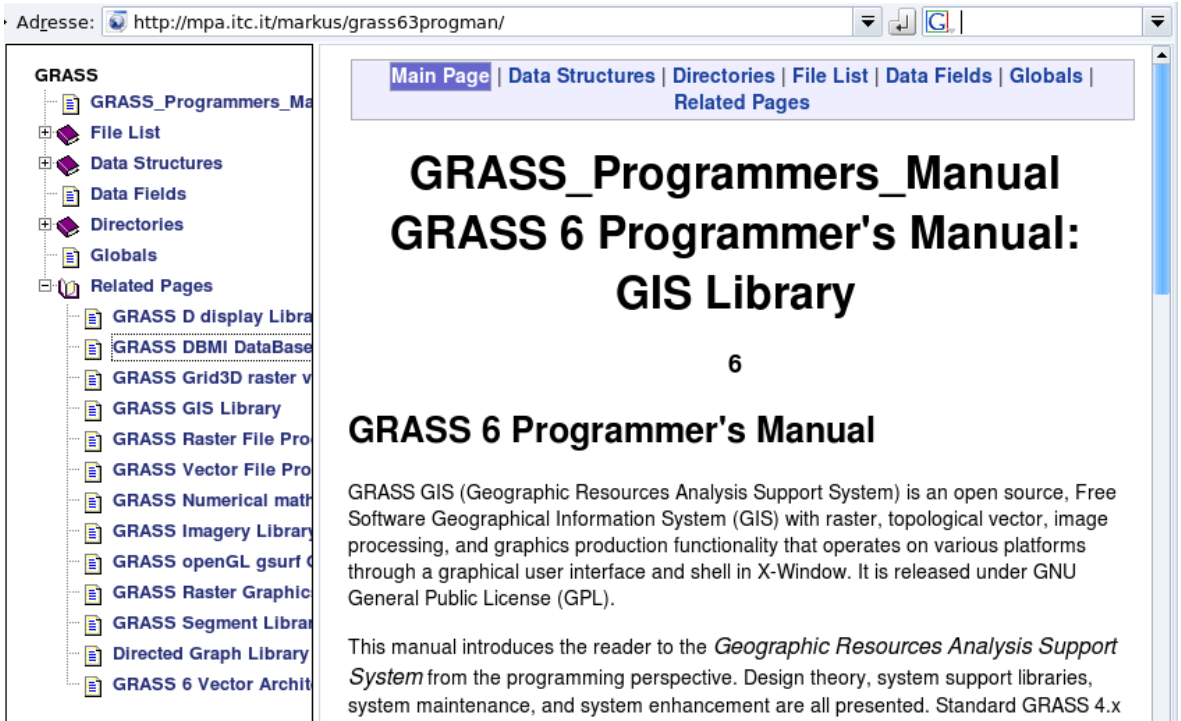
GRASS Documentation



GDF Hannover bR (2005)
free document, GNU FDL
www.gdf-hannover.de

Neteler/Mitasova
(2002/2004/2007)
Kluwer/Springer
Boston, 420 S.
mpa.itc.it/grassbook2/





The Programmer's Manual is 'doxygen' based, i.e. it is auto-generated from the source code.

Outline

Seminar

- Introduction to the GRASS project
- **Communication structure**
- Code development
- Structure of the development team: be collaborative in the cyberspace
- Legal Issues

The actors

Free Open Source Software (FOSS) Community

Universities and Research Institutes
(e.g. NASA, Minnesota, ITCFBK-irst, Uni TN)

Companies
(e.g. D.M. Solutions Refractions, MRCC)

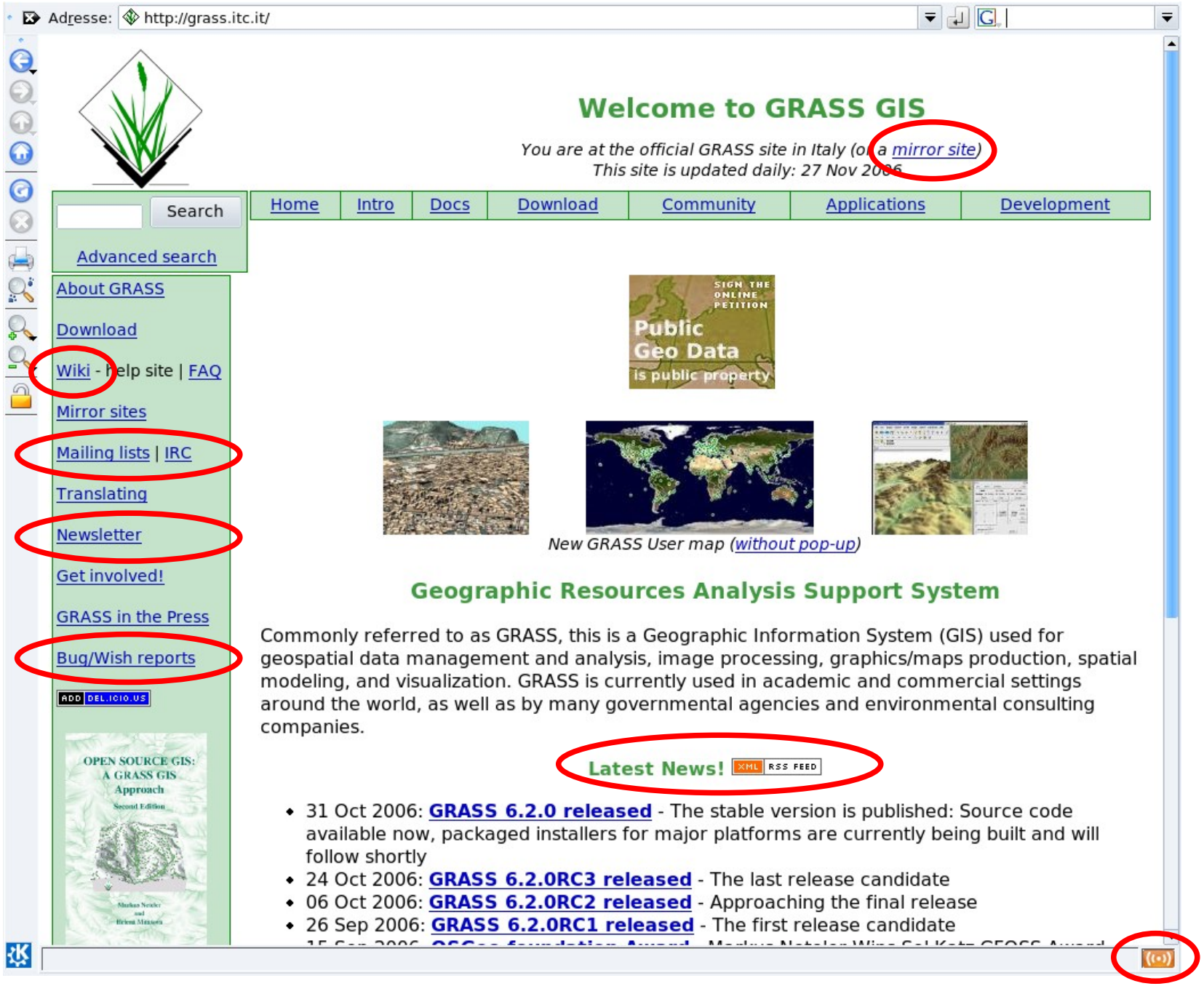
Freelancers
(e.g. F. Warmerdam)

INTERNET
- Web Servers
- Mailing lists
- CVS centralized source code servers
- WIKIs and bulletin boards

Individuals
(often major part of developers and users)

Governments
(e.g. Canada, Japan, Germany, ...)

Communication tools: Project Portal

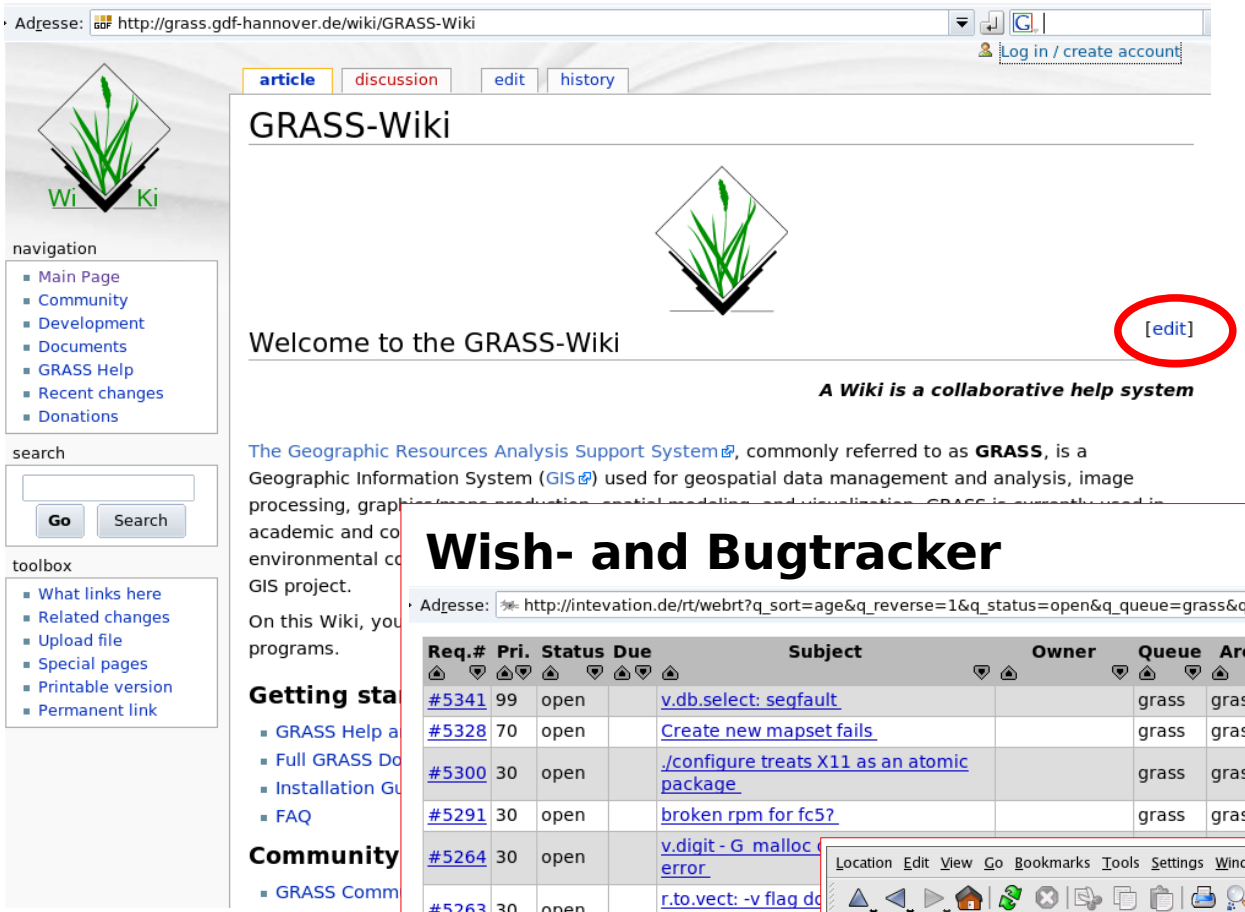


Job automatization: Let the computer do it!

Cronjobs are a life saver!

- Web pages are maintained in CVS and updated via **cron** hourly
- Mirrors sites sync through **rsync**
- Weekly **software snapshots** are generated from CVS
  - source code tarballs
  - binary builts
  - HTML and PDF manual pages
  - local search engine

Communication tools: Wiki and Bugtracker



Wish- and Bugtracker

(old tracker)

Adresse: http://intevation.de/rt/webrt?q_sort=age&q_reverse=1&q_status=open&q_queue=grass&q_area=grass6

Req.#	Pri.	Status	Due	Subject	Owner	Queue	Area	Requestor	Age	Last
#5341	99	open		v.db.select: segfault		grass	grass6		13 hr	72 min
#5328	70	open		Create new mapset fails		grass	grass6		2 day	31 hr
#5300	30	open		./configure treats X11 as an atomic package		grass	grass6		10 day	9 day
#5291	30	open		broken rpm for fc5?		grass	grass6		11 day	10 day
#5264	30	open		v.digit - G_malloc error						
#5263	30	open		r.to.vect: -v flag do values						
#5258	30	open		v.patch: crash or E missing in input						
#5257	30	open		Tcl install copy fai						
#5253	70	open		grass-fc4.rpm dep fontserver (xfs)						
#5220	30	open		wingrass: v.in.ogr creates wrong dire						
#5219	30	open		d.vect display=cat layer 1						
#5218	30	open		wingrass: creating startup screen wit fails						

(new tracker)

Location Edit View Go Bookmarks Tools Settings Window Help

Location: http://wald.intevation.org/tracker/?atid=204&group_id=21&func=browse

FORGE

code I | Search | Advanced search | Log In | New Account

Home | My Page | Project Tree | Code Snippets | Project Openings | GRASS

Summary | Tracker | Surveys

code I: Browse | Download .csv

Assignee: Unassigned | State: Open | Order by: (?) ID | Ascending | Quick Browse

ID	Summary	Open Date	Priority	Assigned To	Submitted By
290	Postgres and mysql tests don't test enough combinations	2007-02-13 20:01	2	Nobody	William Kyngesburye
293	configure shouldn't use gdal-config --dep-libs for GDAL linking (GDALLIBS)	2007-02-14 19:36	3	Nobody	William Kyngesburye
295	grass-6.2.1-1suse10.2.x86_64.rpm, failed dependencies: libmysqlclient.so.15	2007-02-15 01:00	3	Nobody	Jörg Pietruska

Outline

Seminar

- Introduction to the GRASS project
- Communication structure
- **Code development**
- Structure of the development team: be collaborative in the cyberspace
- Legal Issues

Changing source code: what happens? (1/2)

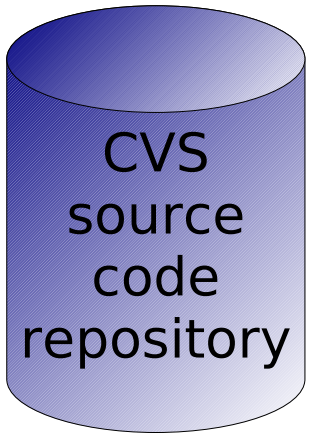
```
tflag->description    = _("Print topology information only");

 if (G_parser(argc,argv))
   exit(EXIT_FAILURE);

 /* open input vector */
 if ((mapset = G_find_vector2 (in_opt->answer, "")) == NULL) {
    G_fatal_error (_("Could not find input map <%s>"), in_opt->answer);
 }
```

Developer changes and enters:

cvs ci -m”i18N macro added” main.c



Germany

Code differences email is auto-generated and sent to “grass-commit” mailing list

Italy

Email notification triggers updated of GRASS Quality Assessment System

Canada

Changing source code: what happens? (2/2)

Email notification triggers updated of GRASS Quality Assessment System

Clone detection is run as well as other quality measures, results sent out

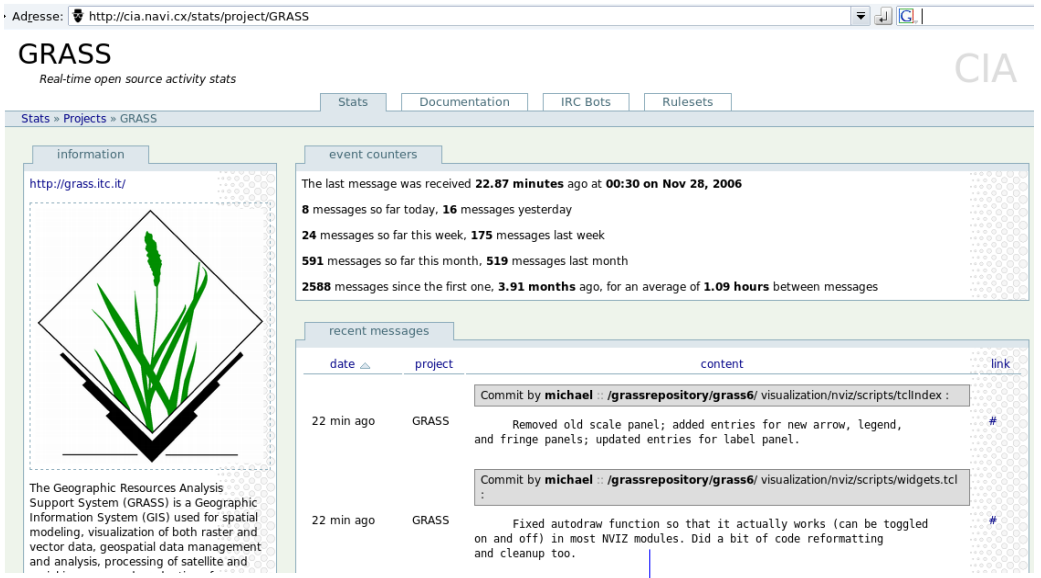
Canada

Code quality email is sent to “grass-qa” mailing list

Italy

CIA open source monitor receives simplified QA message

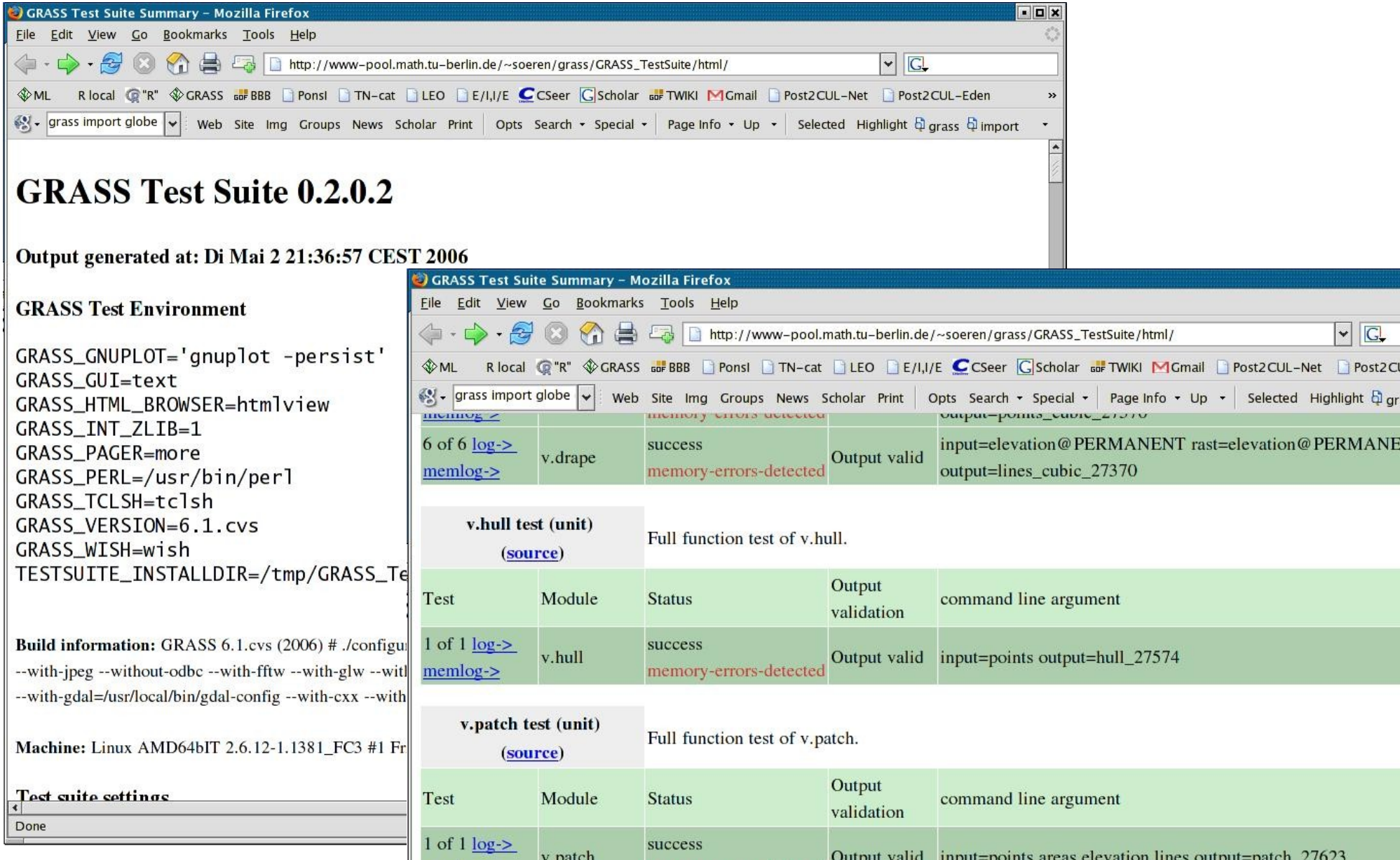
USA?



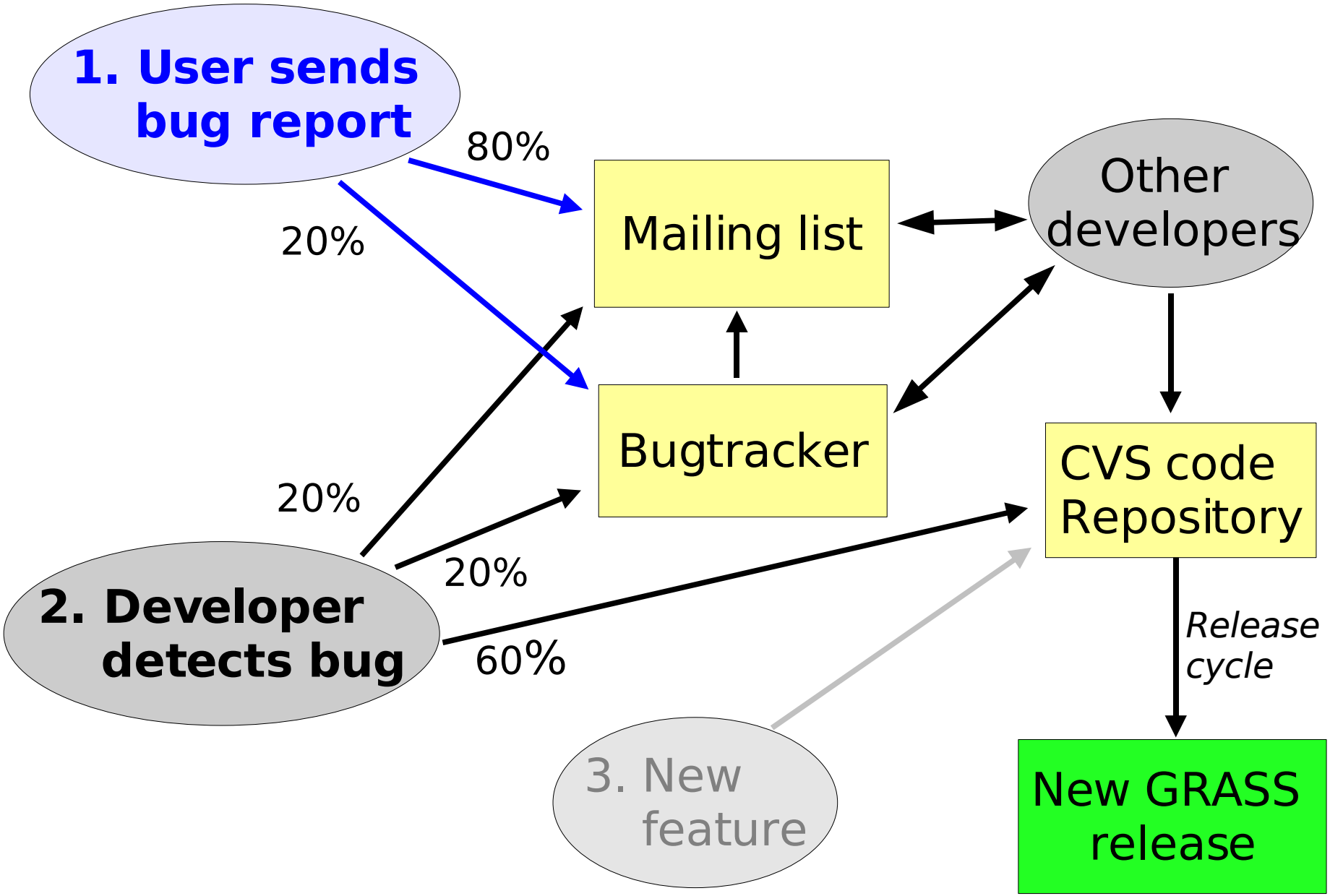
CIA-IRC robot feeds #grass IRC channel on freenode.net

GRASS Quality Assessment I

GRASS Test Suite Project: Automated usage tests on Linux and MS-Windows



Bug reports: Communication Flow



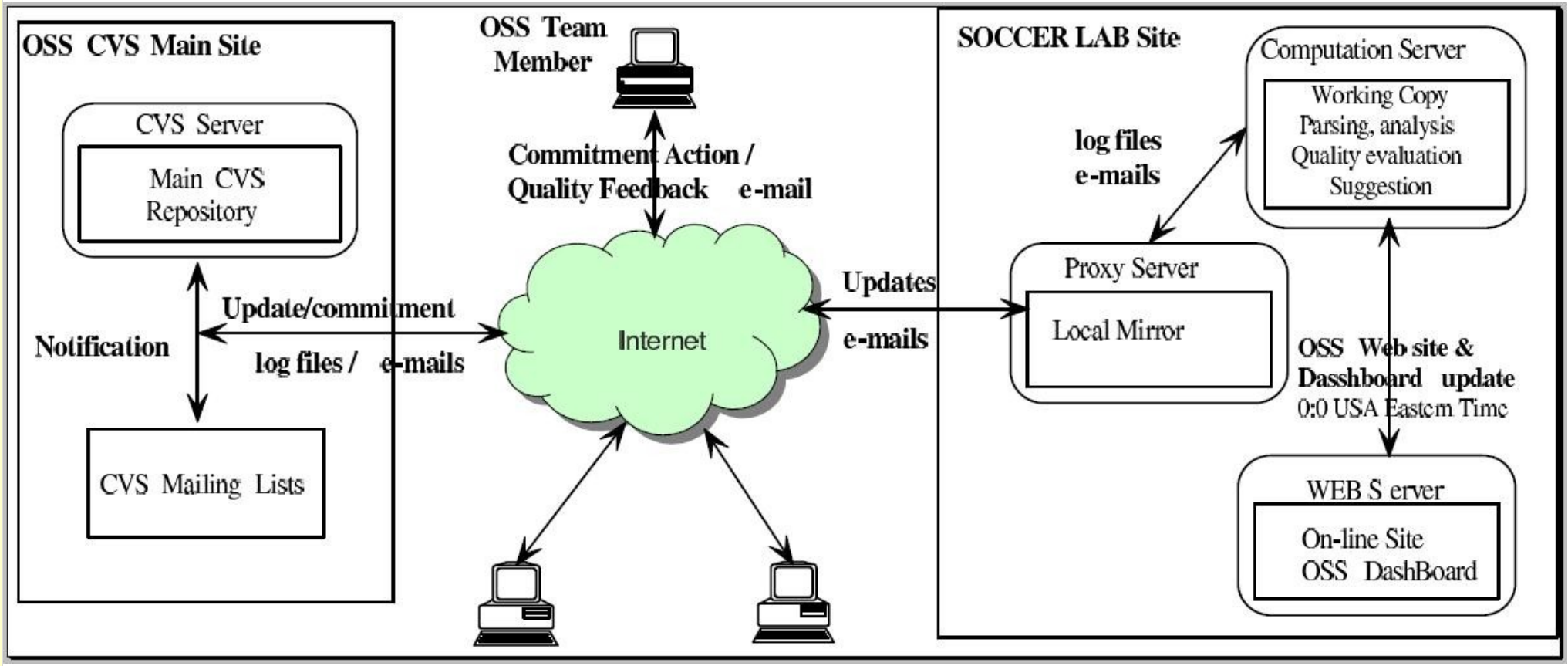
(Percentages are estimated)

GRASS Quality Assessment II

GRASS GIS Software Evolution Project: Software engineering

GRASS Quality Assessment II

Improvement of source code base



Ref.: A feedback based quality assessment to support open source software evolution: the GRASS case study
S. Bouktif, G. Antoniol, E. Merlo, and M. Neteler, ICSM 2006

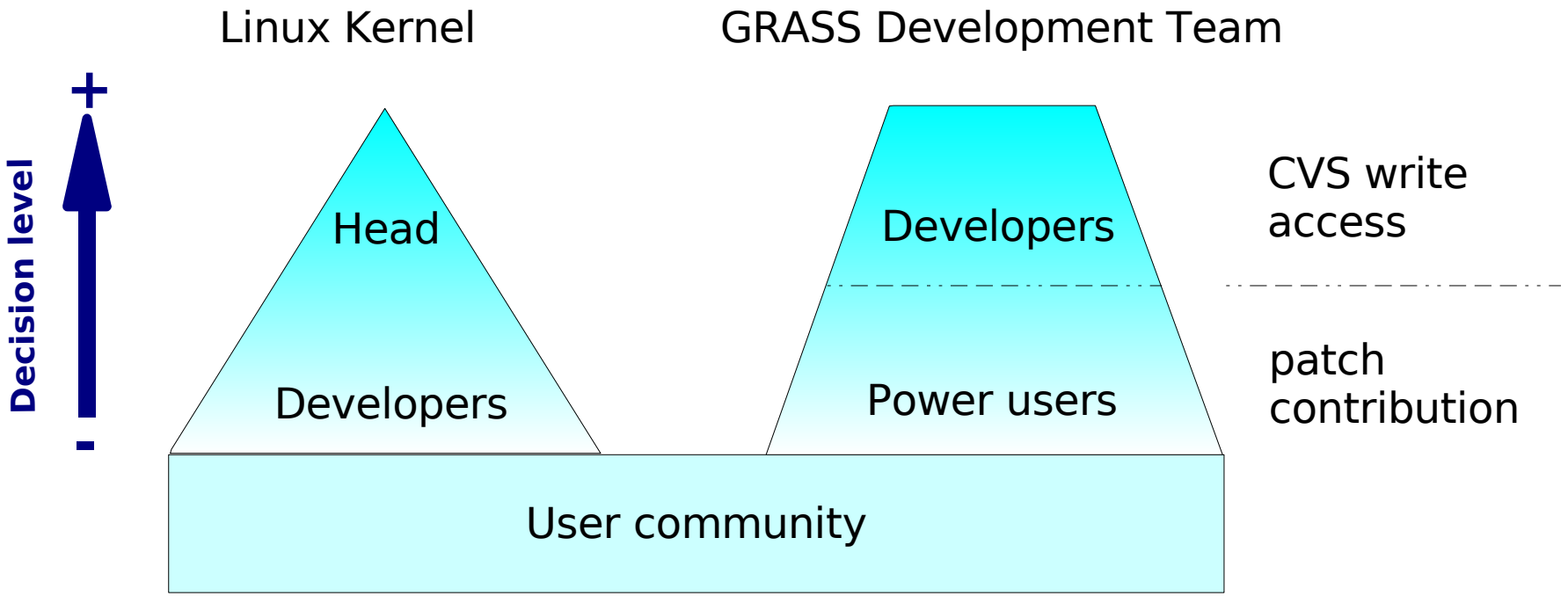
Outline

Seminar

- Introduction to the GRASS project
- Communication structure
- Code development
- **Structure of the development team: be collaborative in the cyberspace**
- Legal Issues

FOSS Software development structures

Organizational structures of development teams



GRASS: No BDFL (Benevolent Dictator For Life)

http://en.wikipedia.org/wiki/Benevolent_Dictator_for_Life

http://producingoss.com/

GRASS Development Team: Structure & “Code habitats”

- Two main types of developers are observed:
  - generalist
  - specialist (majority)

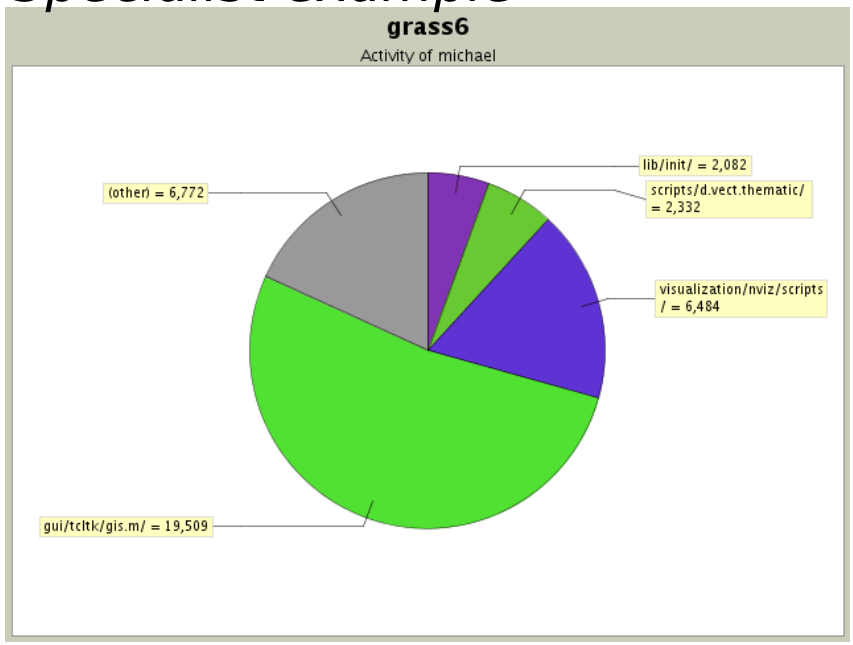
- It appears that many developer assign themselves to a “code habitat”, their area of expertise (in GRASS a selection of libraries or commands which are maintained)
- these “habitats” are often stable over years
- there are also partially abandoned code areas (~ 10% of the code) which are functional but aren't really getting improved

- A very few are experts for code portability (ANSI C etc standards)

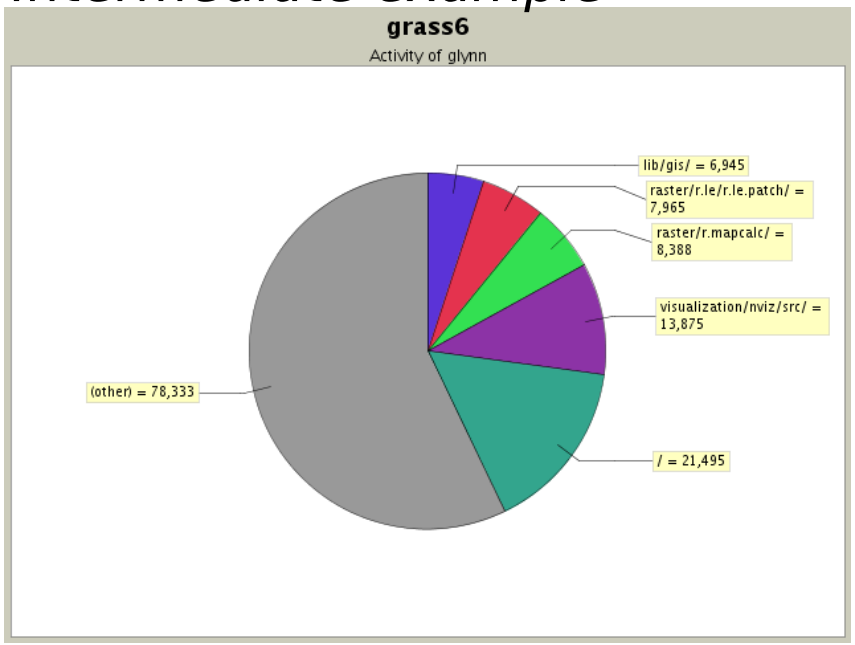
- One “garbage collector” (generalist) fixes lots of *odds 'n ends*

GRASS Development Team: “Code habitats”

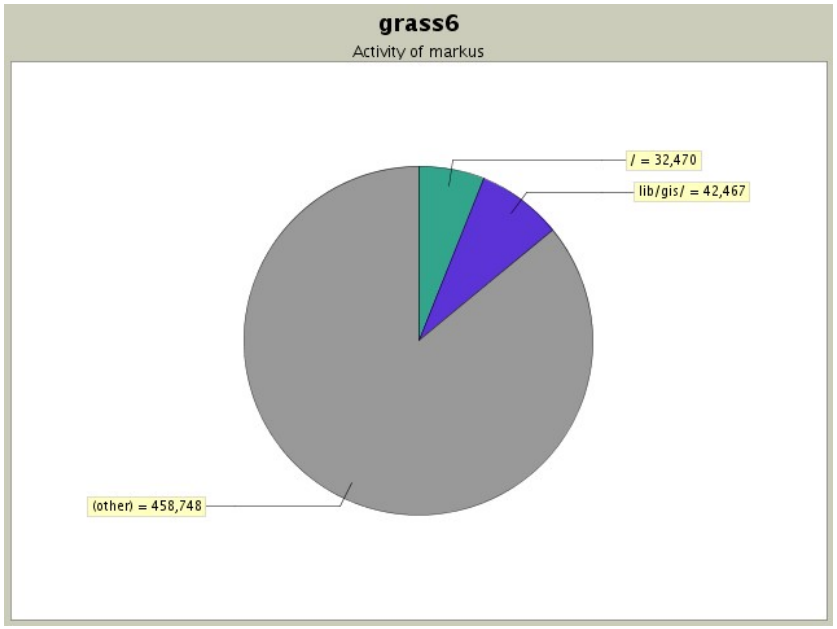
Specialist example



Intermediate example

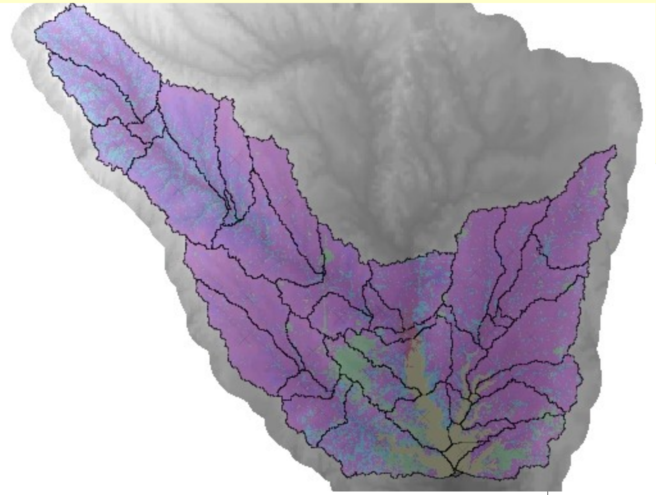


Generalist example



http://web.soccerlab.polymtl.ca/grass-evolution/cvs-stat/overall/module_sizes_glynn.png

Conflicts in the community



Decision making the hard way (1/3)

[GRASS5] Transparency added

[redacted]
Sun, 19 Feb 2006 05:38:08 -0600

- Previous message: [GRASS5] v.out.vtk and r3.out.vtk
- Next message: [GRASS5] Transparency added
- **Messages sorted by:** [date] [thread] [subject] [author]

```
Oops! the attached file is too big.  Try this:
http://geni.ath.cx/grass/transparency.png

---------------
I've added transparency feature to display drivers (XDRIVER and PNG) and d.rast
and d.vect now have transparency= (%) option.  You can find a screenshot
attached in which two rasters and one vector are overlaid.

# opaque drawing
d.rast dem
# 80% transparency, -o is needed not to clip previous drawings
d.rast landuse trans=80 -o
# 90% transparency with blue area fill
d.vect subbasins trans=90 fcolor=blue

Please find attached the png file.

I hope you enjoy this.
```
[redacted]

New feature added...

Conflicts in the community

Decision making the hard way (2/3)

[GRASS5] Transparency added

[redacted]
Sun, 19 Feb 2006 12:48:27 +0100

- Previous message: [GRASS5] Transparency added
- Next message: [GRASS5] Re: [GRASSLIST:10405] Transparency added
- **Messages sorted by:** [date] [thread] [subject] [author]

On Sun, 19 Feb 2006 05:38:08 -0600
[redacted]

> *Oops! the attached file is too big. Try this:*
> *http://geni.ath.cx/grass/transparency.png*
>
> ---
> *I've added transparency feature to display drivers (XDRIVER and PNG)*

Great stuff!

[redacted]

First reactions, but...

Conflicts in the community

Decision making the hard way (3/3)

[GRASS5] Re: [GRASSLIST:10405] Transparency added

[redacted]
Sun, 19 Feb 2006 13:08:47 +0000

- Previous message: [GRASS5] Transparency added
- Next message: [GRASS5] gis.m and d.m gone
- **Messages sorted by:** [date] [thread] [subject] [author]

[redacted] wrote:

> *I've added transparency feature to display drivers (XDRIVER and PNG) and d.rast*
> *and d.vect now have transparency= (%) option.*

Please take a copy of your work, because I'm going to revert these
changes shortly.

--
[redacted]

(later the day an explanation was posted)

An important developer opposes (this code section is “his” habitat)

Decision making

GRASS project:

- rather clear expertises of the developers
- “habitats” can be observed – developers only work on code families
- discussions (even lengthy) via “grass-dev” mailing list [1]
- New GRASS Project Steering Committee (PSC) formed in 2006
- formal voting on “Requests For Comments” (RFCs) but only for CVS access granting and “political” decisions

Other projects:

- similar to GRASS project, BUT:
- RFC voting also for technology changes

[1] http://grass.itc.it/pipermail/grass-dev/

Outline

Seminar

- Introduction to the GRASS project
- Communication structure
- Code development
- Structure of the development team: be collaborative in the cyberspace
- **Legal Issues**

Code vetting

Legal aspects

- License complicance (GRASS: GPL)
- Don't copy from books like “Numerical Receipes in C”
- Ensure that 3rd party contributions are clean
- Employers must agree that worktime is spent

Full transparency and peer review help to minimize the risk.

Apache or OSGeo Foundation

- Incubation phase
- Graduation

http://incubator.apache.org/
http://www.osgeo.org/incubator

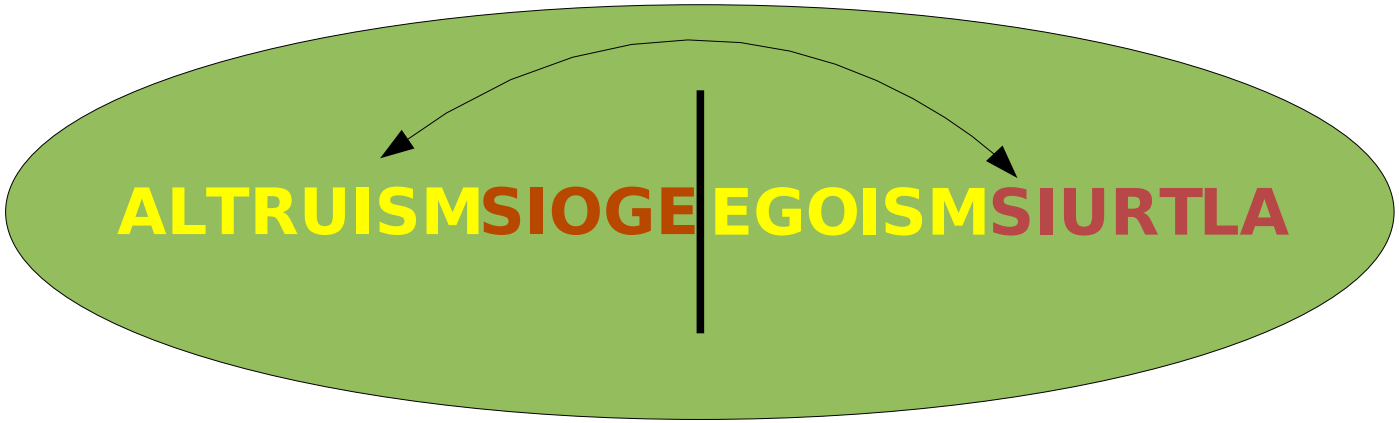
OSGeo Foundation: Founding projects

Founded 4th February 2006, Chicago

http://www.osgeo.org

Why does a developer contribute to Free Software?

I will help others (because) they will help me

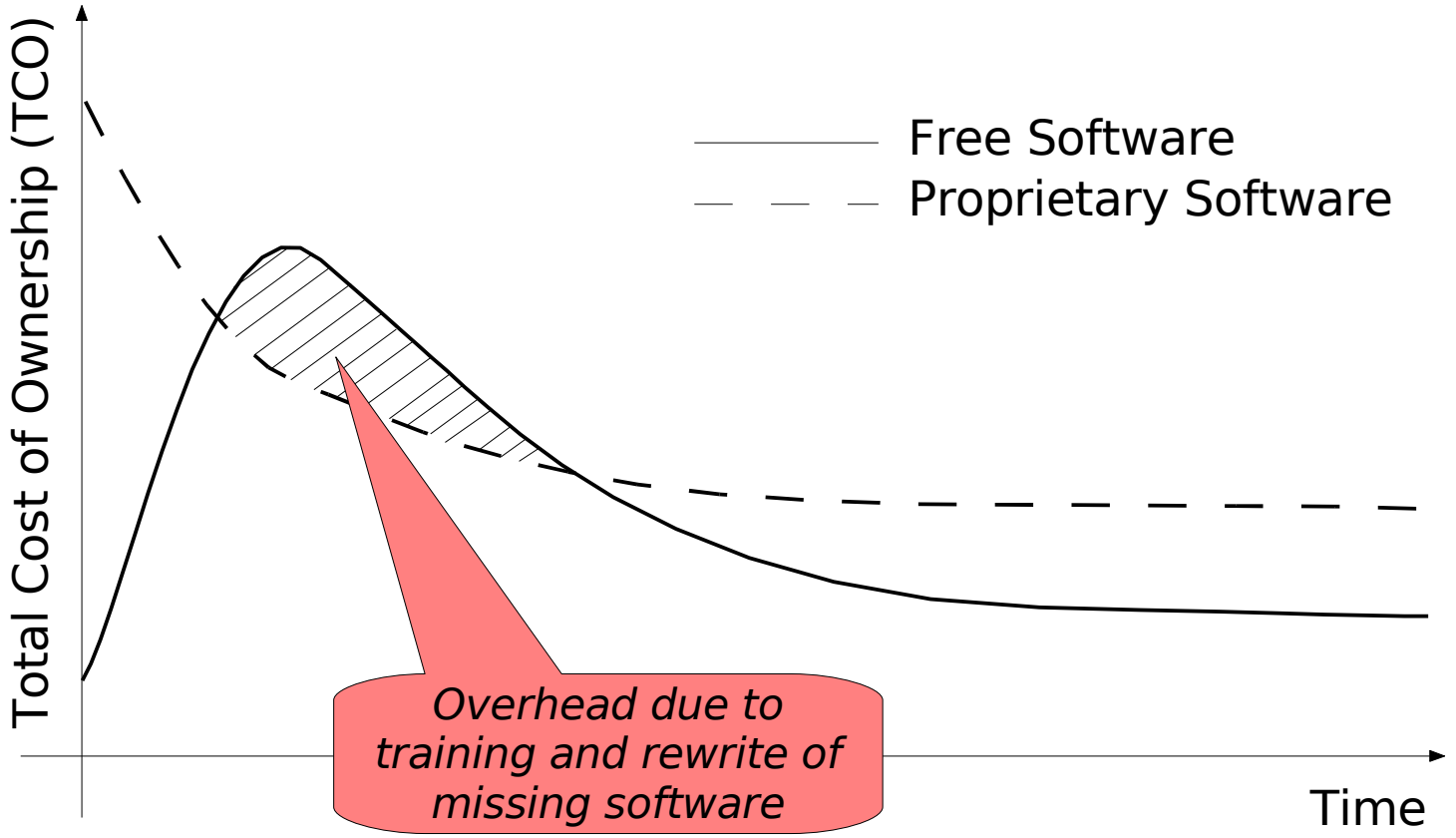


Everyone is expert of only a limited area...
...ask the expert if you don't know!

The driving force behind FOSS development
is **meritocracy**.

License of this document

Software operating costs (customer)



B. Reiter 2004
after Wheeler 2004