



Free/Libre and Open Source Software Metrics

The project addresses the need for more factual information about libre (free, open source) software with an in-depth empirical analysis of thousands of projects, performed by researchers and experts from Spain, The Netherlands, Austria, Greece, Belgium and Italy.

At A Glance: FLOSSMETRICS

Full Title: Free/Libre/Open Source Metrics and Benchmarking

Project Coordinator:

Jesus M. Gonzalez-Barahona

Universidad Rey Juan Carlos

Email: jgb@gsyc.es

<http://flossmetrics.org>

Partners:

Universidad Rey Juan Carlos (ES)

University of Maastricht (NL)

Wirtschaftsuniversitaet Wien (AT)

Aristotle University of Thessaloniki (GR)

Conecta s.r.l (IT)

ZEA Partners (BE)

Philips Medical Systems Nederland B.V.
(NL)

Website:

<http://flossmetrics.org>

Database website:

<http://melquiades.flossmetrics.org>

Duration:

Sept. 2006 – Feb. 2009

Project funding (EC/total):

€583,800/€585,180

Further Information

- **IST Programme:** 2.5.5 Software and Services (FP-6-2005-IST-5)

- DG Information Society & Media
Unit D3 Software Technologies
INFSO-ST@cec.eu.int

- **Europe's Information Society:
Thematic Portal:**

http://europa.eu.int/information_society/

Industry, SMEs, public administrations and individuals are increasingly relying on libre (free, open source) software as a competitive advantage in the globalisation, service-oriented software economy. Because of that, detailed, reliable and complete information about libre software is needed, and specifically about its development process, its productivity and the quality of its results. It is important to know how to benchmark individual projects against the general level, and how to learn from, and adapt, the methods found in libre software to their own development processes, especially within industry.

FLOSSMETRICS addresses those needs by analysing a large quantity (thousands) of libre software projects, providing detailed quantitative data about them. Several aspects of libre software development (software evolution, human resources coordination, effort estimation, productivity, quality, etc.) will be studied in detail.

The main results of FLOSSMETRICS will be:

- a huge database with factual details about all the studied projects;
- some higher level analysis and studies which will help to understand how libre software is actually developed;
- a sustainable platform for continued, publicly available benchmarking and analysis beyond the lifetime of this project.

With these results, European industry, SMEs, as well as public administrations and individuals will be able to take informed decisions about how to benefit from the competitive advantage of libre software, either as a development process or in the evaluation and choosing of individual software applications. The project methodologies and findings go well beyond libre software with implications for evolution, productivity and development processes in software and services in general.

Thousands of libre software projects will be analysed, providing extensive quantitative data about libre software development



European Commission
Information Society and Media

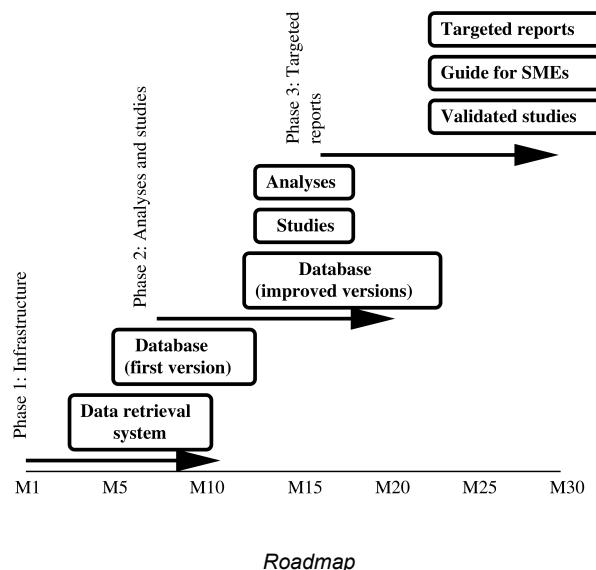
Why FLOSSMETRICS?

In recent years, libre software has developed as a novel form of collaborative production. Since its origin as a collaboration between individual volunteers, it has seen tremendous success, both in terms of the commercial and technical strengths of the produced software itself, but also as a model of organisation and development: open source software is arguably one of the best examples of open, distributed models for production and development that exists today. What is more important, from the point of view of the classical approaches to development methodologies by groups of professionals (and specifically from the point of view of the classical concepts of software engineering), the models used in libre software development are innovative in several ways, to the point that they are only recognized as valid models at all since they have actually produced mature and stable software: any previous "theoretical" analysis would have probably concluded that libre software development was not capable of producing any sustained, useful output.

In this context, FLOSSMETRICS will analyse in depth, from a quantitative point of view, a large number of projects, using mainly publicly available data sources. This analysis will help to better understand the landscape of libre software development, and to obtain factual data about it which can be used to improve libre software development itself (be it done in volunteer or corporate contexts), and to identify interesting practices that could be used in other contexts, but also to obtain indicators and data useful for companies willing to use libre software, or for public administrations interested in its promotion or adoption. In addition, a huge database with quantitative data about thousands of libre software projects will be made available for the use of other research groups, what hopefully will act as a motivator to increase the empirical research on libre software development (and on software development in general).

Roadmap

FLOSSMETRICS is scheduled in three main phases (running partially in parallel). The first one did set up the infrastructure for the project, and the first version of the database with factual data. During the second phase most of the studies and analysis are being performed, and the contents of the database are being enlarged and improved. During the third phase the results of the project will be validated and adapted to the needs of the target communities.



Achievements (Sept. 2008)

During the first two years activities were focused on the design and development of all the infrastructure (retrieval system and database), and the feeding and filling of the database with data from thousands of project repositories. At this point, data is still being retrieved, but also high level analyses and studies are being performed.

The next list summarises the main tasks carried out during this period (Sept. 2006-Sept. 2008):

- **FLOSSMETRICS database:** this database stores quantitative data about libre software projects. Currently, it contains information about more than 1,200 projects, from GNOME, Apache, SourceForge, ObjectWeb and other sites. The retrieved data of these projects are obtained from source code management systems, mailing lists and issue tracking systems. During the next months the analysed FLOSS projects will reach the number of 5,000.
- **Retrieval system:** this software package automates the retrieval (and partially, analysis) of data from public repositories about libre software development. This system is used to feed the FLOSSMETRICS database.

- Guide for SMEs:** this guide (currently in its second version) presents a set of guidelines and suggestions for the adoption of open source software within SMEs, using a ladder model that will guide companies from the initial selection and adoption of FLOSS within the IT infrastructure up to the creation of suitable business models based on libre source software. This guide has had good reviews and has been used into a course of the "Copenhagen Business School".

Its latest version can be found in the next URL: <http://flossmetrics.org/smesguide>

All the software developed in FLOSSMetrics project is hosted in the Morfeo forge and publicly available at:

<http://forge.morfeo-project.org/projects/libresoft-tools/>

Dissemination

The usability of the results of the project (datasets and studies) will be targeted to several different users: SMEs developing or using libre software (or even interested in it), industrial players developing libre software, and the libre software community at large. Based on the feedback obtained in these contexts, a complete exploitation strategy will also be designed.

Dissemination to these communities will be performed using the project website, specific presentations at conferences, and by organizing a series of workshops. Wide impact of the results will be supported by using open access licenses for all output documents.

The data is also expected to be useful for the scientific community, which could use it for their research lines, thus helping to improve the general understanding of libre software development.

During these two years, a set of dissemination tasks have been performed including talks, conferences or flyers. During the last year, three specific workshops were organised, in coordination with other projects:

- Research Room @ FOSDEM, "Libre Software Communities Meet Research Community" workshop** – Brussels (Belgium), February 23th to 24th, 2008
- Libre Software Research Room @ LinuxTag 2008** – Berlin (Germany), May 29th, 2008
- Workshop on Public Data about Software Development @ Conference on Open Source Software** – Milan (Italy), September 10th 2008.

Melquiades website

As it has been mentioned along this brochure, one of the main FLOSSMETRICS targets is to make publicly available the database and the results obtained from the analysis of FLOSS projects. These data are now accessible through Melquiades website (<http://melquiades.flossmetrics.org>).

The screenshot shows the Melquiades Data website. At the top, there's a search bar with the placeholder "Search Example: Evince" and a navigation menu with links to Main, News, Projects, Research, Contact, and About us. Below the header, a message says "Melquiades is the web interface to the data provided by the FLOSSMetrics project". A "Welcome to Melquiades!" section introduces the project. Two "Featured Project" boxes are shown: one for "Evince" (a document viewer) and another for "Tomcat" (a Java-based web server). On the right, there are "Latest News" and "Recently Added Projects" sections, each with a list of items. At the bottom, there's a link to the RSS feed and other site links.

Melquiades Website

The Melquiades website contains full information and metrics about libre software development from more than 1,200 projects, and growing. This information is available in different formats such as database dumps, graphs or even animations.

This screenshot shows a detailed view of repository analyses for the "Evince - The GNOME Document Viewer" project. It includes sections for "SCM repository analyses", "MLS repository analyses", and "TRK repository analyses". Each section lists tools, their types, dates, and download links. For example, under SCM repository analyses, there are entries for "cvsanaly2" and "cvsanay". Under MLS repository analyses, there are entries for "mistats". The bottom section, TRK repository analyses, indicates "No available analyses".

Evince in Melquiades website

Impact in the software domain

The impact of the project is expected to be large in the libre software development realm (and in the whole software development landscape). FLOSSMETRICS will produce the most complete and detailed view of the current landscape of libre software, providing not only a static snapshot of how projects are performing now, but also historical information about the last ten years of libre software development.

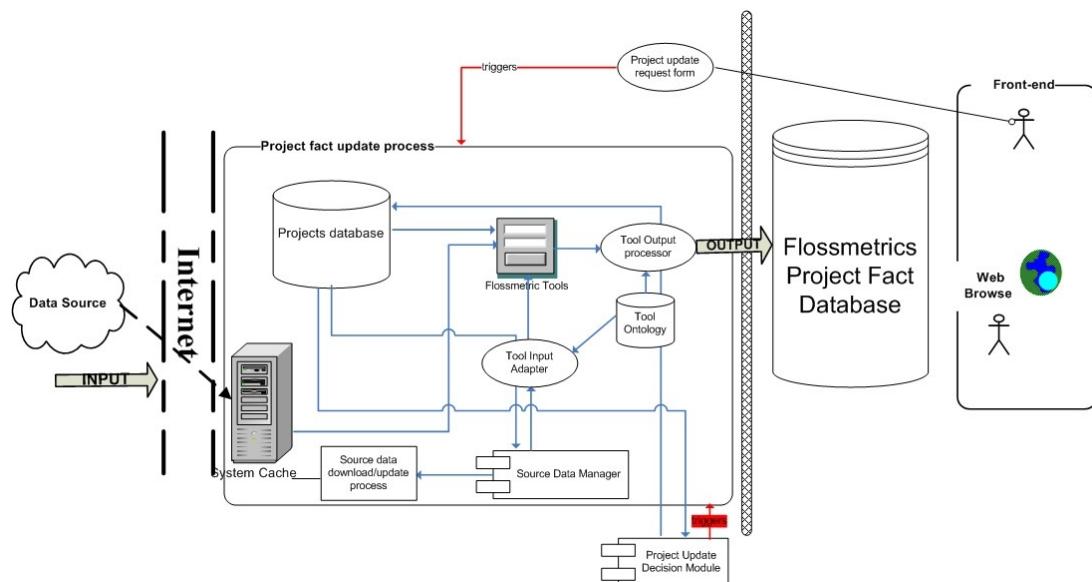
Database Status (September 2008)

- Number of projects: 1200
- Number of analysed repositories: 1100
- Number of analysed mailing lists: 1000
- Number of analysed issue trackers: 2500

Source Code Metrics

Language	Metrics					
	Size			Complexity		
Language	LC	SLOC	#Comments	# Functions	McCabe	Halstead (N,V,D,E)
C	Y	Y	Y	Y	Y	Y
C++	Y	Y	Y	Y	Y	N
Java	Y	Y	Y	Y	Y	N
Perl	Y	Y	Y	Y	N	N
Python	Y	Y	Y	Y	Y	N
Other	Y	Y	N	N	N	N

<http://flossmetrics.org>



Design of the Retrieval System