

**UNITED NATIONS
STATISTICAL COMMISSION AND
ECONOMIC COMMISSION FOR EUROPE**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Electronic Raw Data Reporting
(Geneva, 6-8 November 2006)

Open Source Software in Electronic Raw Data Reporting

Leonhard Maqua, Eurostat

1. MANAGEMENT SUMMARY

Open Source Software is of special interest for Statistical Institutions in several aspects:

- as a source of software products to be used in the statistical process;
- as a way of co-operation with other statistical administrations in the creation of common software modules
- as a way of distributing software developed in-house to customers (especially for electronic reporting)

This paper briefly introduces the open source concept (chapter 2) and presents sources for OSS in the public sector (chapter 3) and discusses legal aspects (chapter 4) and the business model of OSS (chapter 5).

2. WHAT IS OSS?

There are many successful OSS projects. Three prominent examples are:

- Apache, which runs over 70% of the world's web servers.
- BIND, the software that provides the domain name service for the entire Internet.
- Linux, the operating system with the widest range of scalability from embedded devices to supercomputers.

One of the instigators of OSS was Richard Stallman; he started the Free Software Foundation (FSF) and the GNU (GNU's Not Unix) project as early as 1984. Some consider the development of the Internet (or parts of it, e.g. BIND) and Unix (or certain flavours of it) as open source developments that took place even before FSF and GNU. Today's most quoted definition of OSS, however, was written in 1997 by

Bruce Perens who founded the Open Source Initiative (OSI). This OSS definition is known as the Open Source Definition (OSD).

OSS is not public-domain software and is not freeware. Public-domain means that the author surrenders his copyright rights. Freeware does not give modification or redistribution rights to the user. OSS, however, is copyrighted and distributed under a licence which gives the licensee a great amount of freedom in the area of further development (modifications, enhancements, localisation, peripherals, integration, bug fixes and re-distribution).

The benefits of OSS include:

- Software of common interest is made available free of charge to others – expanding the area of development and reducing overall development costs.
- Source code adaptations (e.g. localisation or migration to other platforms) and improvements (e.g. bug fixes or additional functionality) can be made by every user – and reported back to the source code owner who may integrate them into the original code.
- The source code owner can act as the focal point of a group with a common interest – this is for example of interest in the case of EU and ESS where the European Community / European Commission / Eurostat could play this role.

These are examples only, there are more benefits. Various studies show the principal advantages of OSS, being:

- Interoperability, which is one of the main strengths of OSS. However integration with proprietary documents and file formats can be problematic.
- Source code availability.
- Security - complete source code is available (no secret back-doors).
- Quality of Open Source software (errors are usually corrected faster).
- Costs. However, although acquisition costs are generally low, other costs (migration, training, support...) must be considered.
- Stability - no imposed migration to new versions.
- Independence from dominant suppliers - support may be obtained by non-discriminatory Open Calls for Tender.

3. THE ONE-STOP-POINT FOR OSS FOR STATISTICAL ADMINISTRATIONS

In the framework of Eurostat's X-DIS project (financed by the IDANC programme of the European Commission) and in co-operation with the IDABC unit in DG Enterprises and Industry, a CIRCA IG "OSS and statistics" (see picture below¹) has

¹ URL is http://forum.europa.eu.int/irc/dsis/oss/info/data/en/home_page.html

been developed. It is intended to act as a first source of information for all people in statistical administrations interested in using or writing OSS.

The screenshot shows a web page titled "Open Source Software (OSS) and Statistics". At the top, there is a navigation bar with "INFORMATION", "SEARCH", and "HELP" tabs. Below this, there are links for "DSIS:OSS and Statistics" and "Sign in". The page content includes a sidebar menu on the left with items such as "Home", "Introduction to OSS", "Information on the legal aspects of OSS in the public sector", "General OSS resources", "OSS in administrations", "Statistical OSS", "SDMX and other XML and UML resources", and "OSS Glossary". The main content area features a header "OSS and Statistics" with a logo, followed by the title "Open Source Software and Statistics". Below the title, there is a paragraph of text: "There is currently a great interest in Open Source Software (OSS) at public administrations. Several studies have proven the specific benefits of OSS for the public sector. To support the European Statistical System in using OSS, Eurostat is has prepared this CIRCA interest group on OSS." Below this, it says "The interest group is open for your contributions; please contact Eurostat's project officer [Leonhard Maqua](#)". At the bottom of the page, there is a search bar with the text "Find in this group" and a "Go" button. The footer contains navigation links: "Subscription And Contact Information", "Comments", "IG Home Page", and "Site Map".

When developing this CIRCA IG, we found, that there is a growing list of statistical applications that are OSS. These are listed in the above mentioned CIRCA IG, a short summary can be found in the following lists:

General Statistical OSS:

The *R Project* is an OSS tool under GNU for statistical analysis

GGobi is an OSS data visualization system for viewing high-dimensional data

GNUPlot is a portable command-line driven interactive data and function plotting utility, in use by INSEE

Dashboard is a free software written by Jochen Jesinghaus of the European Commission's Joint Research Centre to visualise complex indices and their relations

iPlots is a package for the R-project which provides high interaction statistical graphics, written in Java.

Mondrian is an advanced statistical data-visualization system written in JAVA, with the main emphasis on for Categorical Data, Geographical Data and Large Datasets

Data Transmission Tools:

GENEDI is a tool to create GESMES messages

The freeware *POLYVAL* is a popular GESMES and GESMES/TS tool

Another useful GESMES tool is the *ECB Checker*.

The *ECB GESMES/TS Generator* is a simple Excel application to generate GESMES/TS

Software for Primary Data Collection:

The *RDRMES Toolbox* was created on behalf of EEG6 WG 4

Chiba is an Open Source Java Implementation of the W3C XForms standard

The orientation of the IDABC programme towards OSS also results in all applications developed within the X-DIS project to be published as OSS; especially a number of SDMX related tools has been or is being developed. As well, software produced in statistical research project is more and more published as OSS.

4. LEGAL ASPECTS

The Open Source Definition mentioned in chapter 2 is not, in itself, a software licence. The most popular examples of OSS licences are the GNU General Public Licence (GPL) and the Berkeley System Distribution (BSD) licence, but there are more. An OSS licence protects the copyright of the software author, but gives the users more rights than they get with non-OSS products. These rights include, for example, free re-distribution and the right to modify the source code.

Although there are numerous licences, they all fall within the category of either Permissive or Restrictive Licences. The main difference between these licences is that whereas a Permissive Licence, such as BSD, allows redistribution of the software as part of a commercial package, Restrictive Licences, such as GNU GPL, prohibit users from distributing any program, derived “in whole or in part” from the software, other than under the original terms of the open source software. The significance here is that if commercial developers incorporate an open source product, distributed under a Restrictive Licence, into their products, then they would lose the rights to distribute under vendors’ licences, thereby surrendering the intellectual property rights and commercial viability of their product.

The EUPL (European Union Public Licence) developed by DG Enterprises and Industry, has been approved by the European Commission's Legal Service, and will soon be formally adopted by the European Commission as the preferable OSS licence scheme to be used by Commission services. Opposite to the mostly US-based popular OSS licences like GPL, the EUPL is in accordance with European law; it is therefore as well recommended for use by other administrations and software publishers in Europe.

The draft text of the EUPL is available in the CIRCA IG "OSS and statistics", the final text adopted by the Commission will be uploaded at the same place.

Independent from the licensing issues, there is one major advantage of OSS for public administrations: It allows for really non-discriminatory Calls for Tender. As

all companies have the same free access to the product, they can participate without legal restrictions in tendering processes both on services and on further development.

5. OSS AS A BUSINESS MODEL FOR CO-OPERATION IN SOFTWARE DEVELOPMENT

Although OSS software is normally available without costs, this does not stop companies (or administrations) from making profit with OSS. The business model well adopted in the software industry is to charge for services around the project (consulting, training, hosting, installation etc.).

Second, the OSS model is a good way to distribute software from an NSI to its respondents; OSS software is more trusted than closed modules, and the licensing reduces the liability risks, which are legally inherent to any piece of software sold or distributed for free in Europe.

Finally, OSS is the ideal model for co-operation projects between NSIs. OSS platforms like SourceForge support the shared development of code and documentation. No partner would need to commit himself for a longer period, so that it is easy to get an overview over the financial implications of such a co-operation.

In order to facilitate the publication of statistical software as OSS, guidelines and a questionnaire have been produced in the framework of the above-mentioned X-DIS project. They will be made available on the CIRCA IG on OSS and Statistics..