Distr.
GENERAL

CES/AC.71/2004/15
9 March 2004

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (iii): Open source and software consortia in statistics

EUROSTAT AND OPEN SOURCE

Invited Paper

Submitted by EUROSTAT [1]

I. CONTEXT

1. In October 2002, following discussions at the CoRD (Collection of Raw Data) Task Force, and also at the STNE (Statistics, Telematics Networks & EDI) Working Group, it was decided to prepare a discussion paper on open source in statistics (OSS).

2. A concrete proposal to be discussed was “the establishment of a central OSS group within the European Statistical System (ESS) and the creation of a central repository of statistical OSS”.

3. This paper briefly reports on the open source concept in section II, followed by four case studies at Eurostat (section III) and some Commission studies and activities in relation to OSS in the public sector (section IV). Some general conclusions are at section V followed by some suggested discussion issues at section VI.

II. WHAT IS OSS?

4. There are many successful OSS projects. Three prominent examples are:
- Apache, which runs over 50% of the world’s web servers.
- BIND, the software that provides the domain name service for the entire Internet.
- Linux, the first practical free operating system.

5. 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

[1] Prepared by Leonhard Maqua, Uwe Kunzler and Brendan McAvinue (leonhard.maqua; brendan.mcavinue, uwe.kunzler@cec.eu.int).

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).

6. 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 covered by a license 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).

7. The OSD (see annex 1) is not, in itself, a software license. The most popular examples of OSS licenses are the GNU General Public License (GPL) and the Berkeley System Distribution (BSD) license, but there are more. An OSS license 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.

8. 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.

III. CASE STUDIES

9. Eurostat, like many public administrations and enterprises, are increasingly considering the options offered by Open Source Software. Some Eurostat products, such as the following case studies, have been developed using Open Source architecture. The current policy at Eurostat is that OSS must be considered for all new projects.

III.1 CIRCA

10. CIRCA, the Communication and Information Resource Centre Administrator, is an Internet based groupware tool developed for and owned by the European Community. The European Commission acts as licensor on behalf of the European Community. The CIRCA source code is available to European agencies and national administrations and is in widespread use throughout the Community.

11. Although the architecture of CIRCA is based on open source products such as Linux, Apache and My SQL, the CIRCA license is not compliant with OSD due to a number of restrictions, such as:

- The license is restricted to certain European authorities at national and international level;
- The license is granted explicitly and personally and has to be signed;
- The license is granted for a period of 3 years;
- Commercial use of CIRCA is excluded.

Despite these limitations, there are currently more than 40 licensees in a dozen Member States.

III.2 IDEP/CN8

12. IDEP/CN8, the Intrastat Data Entry Package with the Combined Nomenclature at 8 digit level, is an electronic form for Intrastat declaration, developed and owned by the European Community. Currently there are approximately 40,000 users of the software in nine Member States. Until December 2003, under the EDICOM project, Eurostat developed and maintained the software, and distributed annual versions to member States.

13. Following a decision to cease centralised support for IDEP/CN8 after 2003, the decision at Eurostat is that the copyright to the source code will be retained by Eurostat, whilst legal ownership of the code will be transferred, under legal agreement, to Member States who will be free to modify and further develop the software according to their requirements. However, since the transfer is limited to certain bodies and there are restrictions regarding distribution of the software by each MS, IDEP/CN8 is not fully OSD compliant.

III.3 STIPES

14. STIPES (Statistical Inquiries from Popular European Software) is an IDA funded Eurostat project in the framework of SERT. The final product of STIPES, completed in February 2004, is a generic transformation software that will convert data files from one format to another. STIPES has been developed using open source products and is being considered by Eurostat as a possible pilot case within an open source strategy.

III.4 GENEDI

15. GENEDI is a tool for enabling conversion of statistical tables (in CSV format) into the EDIFACT GESMES (GEneric Statistical MESsage) format. It has been developed using the open source language PERL and runs on any system with a PERL interpreter (MS Windows NT, 9x, 2000, XP; MAC OS, UNIX, LINUX, etc.). GENEDI is made available as freeware.

IV. OSS WITHIN THE PUBLIC SECTOR

16. The case studies in section III concerned development of products and tools which could then possibly be released as OSS. However, a much bigger issue, particularly for public administrations and large offices, is the utilisation of existing Open Source products and the migration to Open Source systems. A number of studies have been carried out by the EU Commission into the possible use of OSS by public administrations:

- Study on the use of Open Source in Europe (June 2001)
- Pooling Open Source Software (June 2002)
- IDA OSS Migration Guidelines (October 2003)

The reports on these studies can be found at: http://europa.eu.int/ISPO/ida (Open Source Observatory / Resources / EU Publications).

IV.1 Main points of studies

Benefits

17. The benefits are:

- 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.
- Costs. However, although acquisition costs are generally low, other costs (migration, training, support etc.) 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.

Drawbacks

18. These are particularly related to organisations, such as the Public Sector, with large ICT infrastructures:

- The existence of a high dependence on the MS/Windows Office Suite, including internal standardisation on servers and desktops. The introduction of heterogeneous components would cause problems in the areas of support, interoperability and data migration.
- Existing long term IT contracts;
- Human resources (with requisite technical skills);
- Lack of pre-installed systems;
- Lack of accountability;
- Hardware and software interoperability (with proprietary products).

19. A proposed IDA project at DG ENTR, **"Encouraging Good Practice in the use of OSS in Public Administrations"**, is aimed at producing an information base on the use of OSS within European Public Administrations. The objective is to set up a focal point which will:

- Give an overview of OSS usage by European public administrations.
- Create an inventory of existing eGovernment applications which may benefit other administrations.
- Provide technical or other advice, to IT policy makers, on OSS-related issues.

Summary

20. Open Source presents *a realistic, technical and economical alternative* to dependence on dominant systems. For large organisations in particular, such as the Public Sector, the long term benefits need to be balanced against the immediate implications for technical and human resources areas.

V. GENERAL CONCLUSIONS ON OSS

21. The general conclusions are the following:

- OSS is feasible.
- OSS is beneficial to Eurostat and other public administrations.
- STIPES should be OSS and could be a test case for the ESS. CIRCA and IDEP/CN8 could be.
- Migration to OSS within the office is feasible, but should be systematic, avoiding a "clean sweep" of existing systems.
- If - for certain legal reasons (especially fears of "unfair competition") - publishing of software developed by an administration under a "real" OSS scheme is not feasible, one should consider a CIRCA-type license.
- A big advantage of OSS for public administrations is that it enables non-discriminatory public procurement.

VI. MATTERS FOR DISCUSSION IN THE EUROPEAN STATISTICAL SYSTEM

22. The following are questions for discussion:

a) Can the European Statistical System agree on a single OSS licensing scheme, or on two alternative schemes to be used? Which ones?

b) Could already existing software be published as OSS and so made available to all partners?
c) Could new software be commonly developed as OSS; and if yes, how to handle
   - co-ordination,
   - localisation,
   - support,
   - funding?
d) Should there be a central repository on statistical OSS (maintained by whom), and/or a central statistical OSS co-ordination instance? Who should maintain this?
e) Should there be an IDA project on statistical OSS? What would you expect from such a project?

References

- CIRCA license: (see annex 2)
- GNU and FSF website: http://www.gnu.org, http://www.fsf.org
- GNU General Public License: http://www.gnu.org/copyleft/gpl.html
- IDA : IDA website, including Open Source Observatory. http://europa.eu.int/ISPO/ida
- Open Source Definition (OSD): http://www.opensource.org/docs/definition.php (see annex 1)
- Open Source Initiative (OSI) website: http://www.opensource.org/
- Raymond, Eric S. (1999). The Cathedral and the Bazaar. O'Reilly & Associates, Inc. 1999 Available online: http://www.catb.org/~esr/writings/cathedral-bazaar/
- Stone, Mark, Sam Ockman, Chris DiBona (1999). Open Sources: Voices from the Open Source revolution. O'Reilly & Associates, Inc. 1999 Available online: http://www.oreilly.com/catalog/opensources/book/perens.html

Glossary

BIND	Berkeley Internet Name Domain
BSD	Berkeley System Distribution
CIRCA	Communication and Information Resource Centre Administrator
CNA	Competent National Administration
DNS	Domain Name System
e-Quest	Austrian electronic questionnaire management system
EDI	Electronic Data Interchange
EDICOM	EDI for Commerce
ESS	European Statistical System
FSF	Free Software Foundation
GENEDI	GENeric EDI
GNU	GNU's Not Unix
GPL	General Public License
IDA	Interchange of Data between Administrations
IDEP/CN8	Intrastat Data Entry Package with the Combined Nomenclature at 8 digit level
OSD	Open Source Definition
OSI	Open Source Initiative
OSS	Open Source Software
PERL	Practical Extraction and Report Language
POSS	Pooling Open Source Software
SERT	Statistiques d'Entreprises et Réseaux Télématiques – Business Statistics and Telematic Networks
STIPES	Statistical Inquiries from Popular European Software
XML	eXtensible Markup Language

Annex 1: The Open Source Definition

Version 1.9

Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:

1. Free Redistribution

The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources. The license shall not require a royalty or other fee for such sale.

2. Source Code

The program must include source code, and must allow distribution in source code as well as compiled form. Where some form of a product is not distributed with source code, there must be a well-publicized means of obtaining the source code for no more than a reasonable reproduction cost – preferably, downloading via the Internet without charge. The source code must be the preferred form in which a programmer would modify the program. Deliberately obfuscated source code is not allowed. Intermediate forms such as the output of a preprocessor or translator are not allowed.

3. Derived Works

The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software.

4. Integrity of The Author's Source Code

The license may restrict source-code from being distributed in modified form *only* if the license allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time. The license must explicitly permit distribution of software built from modified source code. The license may require derived works to carry a different name or version number from the original software.

5. No Discrimination Against Persons or Groups

The license must not discriminate against any person or group of persons.

6. No Discrimination Against Fields of Endeavor

The license must not restrict anyone from making use of the program in a specific field of endeavor. For example, it may not restrict the program from being used in a business, or from being used for genetic research.

7. Distribution of License

The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.

8. License Must Not Be Specific to a Product

The rights attached to the program must not depend on the program's being part of a particular software distribution. If the program is extracted from that distribution and used or distributed within the terms of the program's license, all parties to whom the program is redistributed should have the same rights as those that are granted in conjunction with the original software distribution.

9. The License Must Not Restrict Other Software

The license must not place restrictions on other software that is distributed along with the licensed software. For example, the license must not insist that all other programs distributed on the same medium must be open-source software.

10. The License must be technology-neutral

No provision of the license may be predicated on any individual technology or style of interface.

Annex 2: The CIRCA License

1. LICENSE AGREEMENT. This License agreement covers CIRCA, the Communication and Information Resource Centre Administrator. CIRCA is an Internet based groupware tool developed for and owned by the European Community. As used in this Agreement, "Licensor" shall mean European Commission on behalf of the European Community. Licensee may be an agency or any European national administration (member states, accession counties, TACIS countries and EFTA countries). The License is granted explicitly and personally to Licensee. Licensor explicitly reserves the right to grant the license. Licensee shall not assign or otherwise transfer by operation of law or otherwise this Agreement or any rights or obligations herein. The relationship between Licensor and Licensee is that of independent contractors and neither Licensee nor its agents shall have any authority to bind Licensor in any way.

2. LICENSE GRANT. Licensor grants Licensee a non-exclusive and non-transferable license to reproduce and use for internal or external non-commercial purposes the Product CIRCA, provided any copy must contain all of the original proprietary notices. This license does not entitle Licensee to receive from Licensor hard-copy documentation, technical support, telephone assistance, or enhancements or updates to the Product. Licensee may not redistribute the Product unless explicitly authorised by Licensor in writing. The Licensee is granted to make the product available to a subcontractor for customisation purposes only in accordance with paragraph 3; he is responsible for any breach of the license agreement by his subcontractor. If any services are being provided, then such professional services are provided pursuant to the terms of a separate Professional Services Agreement between Licensor or a third party supplier and Licensee. The parties acknowledge that such services are acquired independently of the Product licensed hereunder, and that provision of such services is not essential to the functionality of such Product. Licensee will inform Licensor on request on any subcontracting of support or development, and whether the product has been made available to the subcontractor.

3. MODIFICATIONS. Licensee may customise, modify or create derivative works of the Product or documentation, including translation or localisation, under the following conditions: (i) Licensee may not redistribute, encumber, sell, rent, lease, sublicense, or otherwise transfer rights to the Product or any modifications to it; (ii) Licensee may not remove or alter any trademark, logo, copyright or other proprietary notices, legends, symbols or labels in the Product; (iii) any modifications must be made available to Licensor and may become property of Licensor if Licensor wishes so, Licensor will provide Licensee with contact addresses for the co-ordination.

4. FEES. There is no license fee for the Product. If Licensee wishes to receive the Product on media, there may be a small charge for the media and for shipping and handling.

5. DURATION. The License is granted for a period of 3 years. It can be extended provided that the Licensee requests it and the Licensor agrees in writing. Without prejudice to any other rights, Licensor may terminate this Agreement at any time if Licensee breaches any of its terms and conditions.

6. PROPRIETARY RIGHTS. Title, ownership rights, and intellectual property rights in the Product shall remain in Licensor. Licensee acknowledges such ownership and intellectual property rights and will not take any action to jeopardise, limit or interfere in any manner with Licensor's ownership of or rights with respect to the Product. The Product is protected by copyright and other intellectual property laws and by international treaties. Title and related rights in the content accessed through the Product is the property of the applicable content owner and is protected by applicable law. The license granted under this Agreement gives Licensee no rights to such content.

7. DISCLAIMER OF WARRANTY. The product is provided free of charge, and, therefore, on an "as is" basis, without warranty of any kind, including without limitation the warranties that it is free of defects, merchantable, fit for a particular purpose or non-infringing. The entire risk as to the quality and performance of the product is borne by Licensee. Should the product prove defective in any respect, Licensee and not Licensor or its suppliers assumes the entire cost of any service and repair. In addition, the security mechanisms

implemented by the product have inherent limitations, and Licensee must determine that the product sufficiently meets its requirements. This disclaimer of warranty constitutes an essential part of this agreement. No use of the product is authorised hereunder except under this disclaimer.

8. LIMITATION OF LIABILITY. In no event will Licensor or its suppliers be liable for any indirect, special, incidental or consequential damages arising out of the use of or inability to use the product, including, without limitation, damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses, even if advised of the possibility thereof, and regardless of the legal or equitable theory (contract, tort or otherwise) upon which the claim is based. In any case, Licensor's entire liability under any provision of this agreement shall not exceed in the aggregate the sum of the fees licensee paid for this license, if any. Licensor is not responsible for any liability arising out of content provided by licensee or a third party that is accessed through the product and/or any material linked through such content.

9. ENCRYPTION. If Licensee wishes to use the cryptographic features of the Product, then Licensee may need to obtain and install a signed digital certificate from a certificate authority or a certificate server. Licensee may be charged additional fees for certification services. Licensee is responsible for maintaining the security of the environment in which the Product is used and the integrity of the private key file used with the Product. In addition, the use of digital certificates is subject to the terms specified by the certificate provider, and there are inherent limitations in the capabilities of digital certificates. If Licensee is sending or receiving digital certificates, Licensee is responsible for familiarising itself with and evaluating such terms and limitations.

10. HIGH RISK ACTIVITIES. The Product is not fault-tolerant and is not designed, manufactured or intended for use or resale as on-line control equipment in hazardous environments requiring fail-safe performance, such as in the operation of nuclear facilities, aircraft navigation or communication systems, air traffic control, direct life support machines, or weapons systems, in which the failure of the Product could lead directly to death, personal injury, or severe physical or environmental damage ("High Risk Activities"). Accordingly, Licensor and its suppliers specifically disclaim any express or implied warranty of fitness for High Risk Activities. Licensee agrees that Licensor and its suppliers will not be liable for any claims or damages arising from the use of the Product in such applications.

11. BASE PRODUCTS. The Product may require other software products (e.g. operating system, Internet server platform) developed by third parties installed before it can be used. These software products are not included in this license agreement, and are by no means under the responsibility of Licensor. Any costs concerning these software products have to be paid by Licensee.

12. COMPLETENESS. This Agreement constitutes the entire agreement between the parties concerning the subject matter hereof. This Agreement may be amended only by a writing signed by both parties. If any provision in this Agreement should be held illegal or unenforceable by a court having jurisdiction, such provision shall be modified to the extent necessary to render it enforceable without losing its intent, or severed from this Agreement if no such modification is possible, and other provisions of this Agreement shall remain in full force and effect. A waiver by either party of any term or condition of this Agreement or any breach thereof, in any one instance, shall not waive such term or condition or any subsequent breach thereof.

13. GOVERNING LAW. This Agreement shall be governed by the Law of the Grand-Duchy of Luxembourg. Unless otherwise agreed in writing, all disputes relating to this Agreement (including any dispute relating to intellectual property rights) shall be subject to final and binding judgement rendered by the competent courts of the Grand-Duchy of Luxembourg. This Agreement shall not be governed by the United Nations Convention on Contracts for the International Sale of Goods.

14. LANGUAGE. The controlling language of this Agreement is English. If Licensee has received a translation into another language, it has been provided for Licensee's convenience only.

15. FORCE MAJEURE. Neither party shall be in default or be liable for any delay, failure in performance (excepting the obligation to pay, if any) or interruption of service resulting directly or indirectly from any cause beyond its reasonable control.

16. HEADINGS. The headings to the sections of this Agreement are used for convenience only and shall have no substantive meaning.

17. REFERENCES. Licensor may use Licensee's name in any customer reference list or in any press release issued by Licensor regarding the licensing of the Product and/or provide Licensee's name to third parties.

18. LOCAL LAW. Licensee is responsible for complying with any local laws in its jurisdiction which might impact its right to import, export or use the Product, and Licensee represents that it has complied with any regulations or registration procedures required by applicable law to make this license enforceable.

- - - - -