Distr.
GENERAL

Working Paper No.9 Addendum 1
19 March 2009

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

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

**ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD meeting on management of statistical information systems (MSIS)
(Oslo, 18-20 May 2009)

MSIS TASK FORCE ON SOFTWARE SHARING - FINAL REPORT

Prepared by the Task Force [1]

I. INTRODUCTION

1. This report examines the possibilities of improving the level of cooperation and sharing of software and related solutions between statistical agencies and international institutes. It has been approved by the MSIS (Management of Statistical Information Systems) Steering Group.

2. The 2009 MSIS meeting is asked to consider the proposals of the Task Force, in particular the establishment of a Sharing Advisory Board. If the proposals for the Sharing Advisory Board are accepted, preparatory work will commence and a formal mandate will be sought from the Bureau of the Conference of European Statisticians at its October 2009 meeting. The Bureau has already given a favourable opinion in principle to the creation of the Sharing Advisory Board.

3. Two projects are proposed. The first, to be carried out by an ESSnet (European Statistical System network) proposed by Eurostat, will take care of the more immediate needs like enhancement of the statistical community on the Open Source Observatory and Repository (OSOR) website and proposals for common architectures etc. The results of this work will be open to all countries. The second includes the setting up of a Sharing Advisory Board reporting to the MSIS meeting, to provide strategic direction and long term continuity. In the following paragraphs reasons and requirements for these two projects will be put forward.

A. Background

4. All over the world, statistical organisations - national (NSIs) as well as international - are carrying out very similar tasks related to the production and dissemination of statistics. This should encourage the sharing of solutions, but in fact each organisation is building their own version of each system, thus duplicating the work, despite many efforts to avoid this situation. Although many computation and data manipulation tasks can be done with Common Off The Shelf software, there is a substantial amount of software and formalized knowledge that is specific to the work in NSIs. In the - all too few - cases where a group of organisations have agreed to use one common solution, normally developed by one of them or through a joint project, there is often no mechanism to secure continuity for the future.

[1] Mr Marton Vucsan (Chair - Netherlands), Mr Alistair Hamilton (Australia), Ms Karen Doherty (Canada), Mr Carlo Vaccari (Italy), Mr Rune Gloersen (Norway), Mr Leonhard Maqua (Eurostat), Mr Antonio Consoli (Eurostat), Mr Lars Thygesen (OECD), Mr Valentin Todorov (UNIDO), Mr Steven Vale (Secretariat - UNECE)

5. The MSIS 2008 meeting identified a strong requirement to explore existing and potential mechanisms for the sharing of software and components between statistical agencies, with a view to further formalisation of the existing practices and to encourage joint development work. This requirement has also been recognized by Eurostat, which has already begun to reserve some funding for activities in this area. Given its wide membership and the fact that members are typically responsible for systems architectures of their organisations, it is seen as appropriate that MSIS should be the body to give strategic direction to this initiative and determine the high level requirements of the proposed project to enhance the sharing of software.

6. During the MSIS meeting of 2008 it became clear that software sharing is a complex matter. Not only are there all kinds of short term needs, but it is also clear that, in the long term, the present variety of platforms and architectures will lead to inefficiencies and unnecessarily high costs. A dual approach is therefore needed for a project to solve these kinds of problems.

7. Firstly, in the short term, a number of activities are needed to progress quickly. This will happen in an environment with lots of different platforms, licence models etc. In the current environment it will be nearly impossible to design things like a common architecture or to converge on a model for sharing. However it will be possible to start with enabling initiatives such as a web portal to share information and tools. The aim is to create a culture of exchange of information and knowledge about sharing, open source and licensing etc. This will hopefully satisfy the need for information and provide an easy pathway to tools that can already be shared within the statistical community. We will call this the tactical solution.

8. The proposed ESSnet will be a very useful instrument to pursue this tactical solution. It will be a network of members of the European Statistical System (ESS), working on a common project co-funded by the European Commission (see the appendix on the ESSnet as proposed at the moment to Eurostat's Statistical Program Committee). Although formal membership of an ESSnet (and hence access to funding) would only be open to members of the ESS, non-members of the EU will be able to take advantage of the project infrastructure while contributing and enjoying the results of the projects. To take advantage of the possibility of at least partial funding from Eurostat, proposals for future actions should include justification for an ESSnet.

9. Secondly, in the long term, we need a solution that will cause the number of computing platforms and licence types to be reduced, and systems architectures to converge. This will not happen overnight. It will be a long and tedious job with lots of documents and proposals. Without attention and direction that is sustained over time it will surely fail. The answer to this problem is governance. If set up properly, it will take care of the convergence needed and provide structure and direction for years to come. To make this happen, there will have to be a governing body to oversee the work being done in these areas and provide guidance and direction. Various working parties and other projects or joint ventures between NSIs will then be able to align and to contribute. We will call this the strategic solution.

II. DELIVERABLES

A. Strategic solution

10. At the strategic level, the aim that seems most important is convergence or support of divergent platforms and solutions. Although a lot of work can be done in the short term (a few years) real success will only come through sustained efforts over a longer period of time. Common architectures and platforms will have to be decided upon and continuing governance and guidance will have to be provided. A useful first step towards the convergence of architectures would be the general adoption of the Generic Statistical Business Process Model, which is currently being developed by the METIS working group on statistical metadata (see: http://www1.unece.org/stat/platform/download/attachments/5472434/GSBPM+Dec_08.doc?version=1.

11. In order to bear long lasting fruits, the proposed project will have to create a governing body to provide direction and ensure convergence long after the project itself has finished. For this body the membership, governance, mission statement, terms of reference, working procedures, responsibilities, decision making powers and initial task list will all need to be defined.

12. The establishment of this governing body and development of the strategic tasks should be carried out by the MSIS group, or a taskforce reporting to it.

B. Tactical solution

13. To get immediate results we propose to further enhance the portal function already present at Eurostat and/or create a specific statistical corner of the Open Source Observatory and Repository (OSOR) site, to facilitate sharing by providing technical, procedural and legal information as well as a repository of information and software to be used by the NSIs. The portal will have to support several different models of sharing at least in the short-term.

14. Existing software that was built in previous collaborative projects or given to the community will have to be sustained in an orderly fashion with a maintenance and release schedule and with funding approved by the community. At the moment quite a lot of software is being maintained by MSIS member countries on a voluntary basis without any formal direction. In other cases (e.g. PC-Axis) a consortium of interested agencies maintains and develops the software. The governing body mentioned in the previous section, in consultation with the countries involved, will determine the most appropriate method in each case to ensure continuity.

15. As a starting point for the creation of a common architecture an initial design should be made with considerable allowance for individual differences. The feasibility of the adoption of models such as the Generic Statistical Business Process Model, referred to above, should be considered.

16. The tasks mentioned above are well suited to be carried out by the ESSnet Eurostat is currently proposing in this area. (See Appendix 1; please note that the original proposal also allows for the creation of standard tools. Although not mentioned in this report, it is clear that tools that facilitate the adoption of common architectures or solutions are desirable and should be created whenever possible.)

III. REQUIREMENTS

A. Strategic solution (to be carried out by MSIS)

17. Strategic goals in software sharing and convergence can only successfully be reached through alignment. It is not a technical problem but a managerial one. Another property of strategic directions and goals is that they only have meaning at a higher level of abstraction and management. A substantial amount of time is needed to reach strategic goals. This implies that structures are needed with a certain amount of continuity. A short-term working party or task force will not be sufficient. We propose to create a governing body to look after the strategic interests of shared development convergence and funding of existing initiatives (the current list of existing software).

Governance

Description of governing body

18. The governing body should be divided into a preparatory group (further to be called Sharing Advisory Board) that will create proposals to be approved, and a group of decision makers that will meet at least once a year. Given that it is the group with the widest membership of heads of information technology in statistical agencies, this seems a suitable role for the MSIS group. If necessary, more strategic decisions can then be referred to the parent body, the Conference of European Statisticians, for approval by the heads of national and international statistical agencies. In this way there is sufficient knowledge and time for thorough preparation of proposals and enough decision power to get approval and budgets to get the proposals executed.

19. A concrete proposal is therefore to create a Sharing Advisory Board and to put a procedure in place for decision making through the MSIS group. The Sharing Advisory Board will then oversee the execution and report back to MSIS, and will be considered a sub-group of MSIS.

20. A mission statement and overview of the tasks to be carried out will be written and agreed with the Bureau of the Conference of European Statisticians. The detailed tasks of the advisory board will be determined and reviewed annually by the members of MSIS.

21. Initially the following tasks are envisaged:

- Proposals for maintenance and follow up of the architecture task;
- Proposals for the maintenance schedule and funding of existing software.

Membership procedures

22. The membership of the Sharing Advisory Board has to be decided upon, together with procedures to enable healthy dynamics of the composition of that board. It is proposed that the board consists of five members, elected by the MSIS group for periods of two years during the annual meeting. Members can be re-elected once. The chair is to be chosen by the board from amongst its members.

Operating procedures

23. In order to guarantee long-lasting continuity, the operating procedures of the Sharing Advisory Board have to be written and agreed upon.

Decision making

24. Procedures will have to be written and approved to create a standard periodic cycle of reporting and approval of the work done by the Sharing Advisory Board.

Future models of sharing and convergence

25. The problem with future models lies in the word future. It will be very difficult to create a template or statement of direction on this subject without some form of follow up. We propose that the ESSnet should create a statement of direction in which an initial proposal for one or more shared architectures and platforms is put forward and reviewed by the Sharing Advisory Board. A task for the Sharing Advisory Board will be to supervise the progress on a regular basis. This task also has to provide for follow up in the form of ensuing projects or taskforces etc.

B. Tactical solution (to be carried out by the ESSnet)

26. In the immediate future we think the following products are needed.

Common architecture outlines

27. For a solution to work effectively, we need to think about the creation of a common architecture. This architecture would have to be based on standards, common tools, shared software products, etc. The choice of a common architecture is a strategic task that has to be adequately prepared technically and this preparatory work fits in well with the development of the tactical solution to be provided by the ESSnet. Creating the common architecture will be a challenge given that NSIs have made different choices in the past resulting in more than one platform and more than one approach being used for the IT component of the statistical process. Another factor to be considered is that, in some cases, architectures may also be influenced by national government policies or decisions. However, one may assume that since we are all making similar statistics there may be common ground at the business level.

28. The basic products as seen by us are:

- Business architecture - a common framework to describe the Business Process Model, with the potential for future harmonisation;
- Information Systems Architecture - consisting of data and metadata architecture and overall application architecture on a logical level;
- Technical Architecture - for example, Service Oriented Architecture, technical platforms, use of standards.

Requirements:

29. The requirements mentioned above reflect the three cornerstones for interoperability, i.e. organisational, semantic and technical interoperability. Usually interoperability is aimed at sharing or linking processes across businesses, but we will also need the same approach in order to share software for common business processes. The architecture has to be very flexible and easily adaptable to change. It has to be modular and it must be possible to develop different components independently, so that different parties could contribute to their implementation, while fixing common rules for their integration to assure that they will work together within the physical architecture(s). Different organisations would also then be free to use the modules they really need.

30. The design of the architecture should also include a plan for its implementation, with a series of phases, so that when a component is ready it can be used without the need to wait for the complete architecture to be finished. The architecture has to be generic enough to try to cover the largest number of needs of the NSIs but keeping a degree of simplicity to enable implementations with simple means in a reasonable time period, keeping also in mind that IT evolves very quickly and that an architecture that is too complex could become obsolete even before the end of its implementation.

31. Even if open source software (OSS) would seem to be a natural choice it must be considered that different NSIs have invested heavily in proprietary software solutions and they cannot be expected to change completely in a short period. The architecture should therefore also support proprietary solutions, at least the most common ones, as well as OSS. This is one of the reasons why more than one physical architecture could be needed and in this case a way to move gradually from one to another should be described.

Integration of existing software

32. The ESSnet shall take into account solutions that already exist. Solutions that are already shared in the statistical community shall be investigated for fitting into the technical architecture. In addition, business models shall be discussed with the owners of these applications concerning how their maintenance can be assured in the long term, and how they could be adapted to the proposed architecture, without giving extra burdens to individual NSIs.

C. Tools for co-operation

Portal on shared statistical software

33. Currently, two portals exist for applications shared between administrations: The OSOR website (http://osor.eu/) and the CIRCA[2] group "OSS and Statistics" (http://circa.europa.eu/irc/dsis/oss/info/data/en/home_page.html), on the latter all shared and free resources for statistical administrations are being made available. While OSOR is maintained by the Informatics Directorate of the European Commission, the CIRCA group is maintained by the Statistical Information Technologies unit of Eurostat. With these portals all information about sharing of software, do's and don'ts etc. are made available. The intended users are decision makers and technical staff in NSIs and other institutes. The ESSnet shall cooperate with these portals and contribute the future models and architectures, so that they can be prominently displayed there as they become available.

[2] CIRCA is a European Commission tool providing Internet work spaces to groups engaged on specific topics.

34. To ease the transition to shared software, abundant information about the various legal forms of sharing should be presented and explained. All models of sharing should be represented, to name a few:

- OSS licenses (European Union Public License - EUPL, General Public License - GPL etc.), with the option to offer paid support;
- Barter (e.g. exchange of self-developed software between partners);
- Consortium for joint development;
- Executable ownership (with escrow for source code);
- Free executable with (varying) financial compensation.

35. If possible, information on how to (slowly) move from one model to the chosen common future models should also be present.

Technical

36. The technical area of the information portal should give guidance and white papers about products available and general background information about technical solutions with their advantages and shortcomings etc. This area is intended for technical staff.

Repository

37. A crucial issue in developing, maintaining and sharing software is the communication among all involved parties, possibly separated in time and distance - the users, the developers, the project leaders, the architects, the designers, and the testers. These parties have different tasks in the development life cycle and use different tools for accomplishing these tasks and at the same time have to collaborate with each other. Such environments integrating all necessary tools for management of the entire development life cycle are popular in the open source development world, usually called Collaborative Development Environments and most of them are based on the GForge project (http://gforge.org/).

38. The repository is an environment for hosting projects in such a way that all project deliverables (the code, the documentation, binaries, etc.) are publicly accessible to all interested organizations, which can use the software that has been already developed, and contribute feedback, bugs, ideas and suggestions.

39. The OSOR repository http://forge.osor.eu offering the full GForge functionality (plus subversion (SVN) for the maintenance of the code, and more) is strongly recommended as the collaborative platform for the ESSnet, as well for any other collaboration projects of statistical administrations.

40. The OSOR repository is closely related to the OSOR observatory, which can easily be used as a platform to market the solutions.

Main Features

41. The OSOR repository is a full configured development system supporting standard development tools integrated into one web page. The required features are provided by the well known collaborative development environment GForge, as used in the OSOR, (http://osor.eu/):

- A project web site;
- A source repository with versioning tool allowing the development teams to safely coordinate and track software source code changes with a choice between CVS (Concurrent Versions System) or SVN;
- Tools for communication between members of a development team like discussion forums, bug tracking, support for change and enhancement requests, submission of patches, mailing lists, posting of news, task lists and to do lists;
- Document management with approval queue.

APPENDIX 1: THE ESSNET AS PROPOSED TO EUROSTAT'S STATISTICAL PROGRAMME COMMITTEE

In the document for the May 2008 meeting of the Statistical Programme Committee, the ESSnet is defined as follows:

Main objectives: To support the development and maintenance of statistical IT tools for use by NSIs and other national authorities, within the general framework of a common reference architecture for the data life cycle at national level, and using appropriate common open standards and guidelines such as SDMX. This ESSnet would deal with the elaboration of a common reference architecture for the data life cycle at national level and the identification and development of the IT tools needed to implement such an architecture. All tools would be developed as Open Source Software and would be licensed under the EU Public Licence or another compatible OSS license. They would be designed to be portable for use in contexts other than the common reference architecture. Use of SDMX standards and guidelines would be required in all appropriate cases. Other XML-based standards such as XBRL and IQML would be used where appropriate.

Outputs:

- Process descriptions and mapping;
- Rules for collaborative development;
- Software to be shared.

Possible Partners: NSIs and other national authorities collecting statistical data from businesses and households.

Start Date: First semester 2009, with first results presented to ITDG 2009.

Duration: 9 months with follow up in successive years.

Responsible Eurostat Unit: Units B3+B1 EC.

Financing: 500 K€ (subject to availability in 2009 Financing Decision).