Distr.
GENERAL

Working Paper No.5
7 May 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**

**Meeting on the Management of Statistical Information Systems (MSIS 2009)**
(Oslo, Norway, 18-20 May 2009)

Topic (i): Sharing of software and components

# COmmon Reference Architecture (CORA) ESSnet: history and next activities

**Invited Paper**

Prepared by Carlo Vaccari, ISTAT, Italy

## I. HISTORY

1. Shared activities between National Statistical Institutes on the topic of “Common Reference Model” (or Architecture or Business Process Model or Statistical BPM or ...) started many years ago.

### A. 1999 – CES document on Information Systems Architecture

2. Already in 1999 the Conference of European Statisticians (CES) published a study about “Information Systems Architecture for National and International Statistical Offices – Guidelines and Recommendations” (see http://www.unece.org/stats/documents/information_systems_architecture/1.e.pdf).

3. In the first chapter of the document the CES explains the basic concepts used in the report: different types of statistical processes and statistical data (micro- macro- and meta-), statistical applications and infrastructure, and the flow of data and metadata through the survey process.

4. The second chapter provides an analysis of statistical processes life-cycle (planning, operation and evaluation) for the four major types of Statistical Information Systems (SIS): survey processing, data warehouses, registers and analytical processing systems.

5. In the third chapter an information systems architecture for a statistical organisation is outlined. The architecture is based upon the four types of SIS specified above and is based on a corporate data warehouse which includes five compartments:

- raw data and metadata;
- final observation registers;
- final multidimensional statistics;
- electronic documents;

- global metadata, including registers.

6. Chapter 4 analyses the technical aspects of different information systems architectures, leading to a proposal for a multi-tier network-based information systems architecture that balances the needs for centralization and decentralization in a modern statistical organization.

7. Chapter 5 focuses on the implementation aspects of the proposed IT architecture. Under the conditions of rapid IT development, there must be a realistic plan for implementation which is able to accommodate changes that happen during the implementation process itself.

8. The CES document was very important for many organizations that adopted it in some way: NSIs from Latvia, Croatia, Lithuania, Ireland are using those guidelines; obviously many aspects of the document became obsolete (e.g. technology issues like Web-SOA).

## B. Open source considerations

9. Some years later in a MSIS session devoted to "Open source and software consortia in statistics" (Gloersen - Janson at MSIS-2004: Open Source for Statistics: a Stairway to Heaven?) Statistics Norway wrote: "The **common usage** of software, combined with the **common model** for statistical output databases and to some extent common use of metadata standards have shown to be very useful for the Group. Concrete interfaces have been developed, and there is increasing number of examples where one country develops modules that easily can be reused by other partners."

## C. Technical developments

10. There was an interesting debate in Sofia (MSIS- 2006) where, talking about UK paper "Towards a Service-Oriented Architecture (SOA) for the Statistical Value Chain", many countries expressed the willing of make some kind of shared job in defining a framework in which to co-operate, using a common architecture inside which to develop shared tools using technology like Web Services. In the UK paper was written: "... it is possible that NSIs could agree on a **statistical value chain** that could be at the level of discrete web services. A topic such as "Other Advanced Analyses" could be subdivided usefully into further analysis domains and specific WSs developed for each."

## D. 2007 – BmTS from New Zealand

11. In MSIS 2007 a paper was presented by Statistics New Zealand with the title: "The BmTS: Creating a new business model for a National Statistical Office of the 21st Century". In the paper the Business Model Transformation Strategy (BmTS) is described, i.e. a model which has to provide a solid basis for growth and development with the design and build of a "platform" on which future programmes and projects undertaken by Statistics New Zealand will be founded.

12. In the conclusions, Statistics New Zealand wrote: "Service Oriented Architecture has the ability to support the needs of a Statistical Information Architecture as long as the implementing organisation has completed the **pre-requisites of business process modelling** (i.e. identified and agreed the value chain) and determined the information architecture which will support the business model."

13. From other presentations you can know what happened in MSIS 2008 and how the decision of starting a Sharing Advisory Board was taken.

## II. STARTING AN ESSNET

### A. ESSnet?

14. In “Statistical work programme of the Commission for 2009” (http://epp.eurostat.ec.europa.eu/portal/pls/portal/!PORTAL.wwpob_page.show?_docname=982227.PDF) the ESSnet tool was launched: “Following the needs expressed by DGINS in 2002 to find synergies, harmonization and dissemination of best practices in the European Statistical System, Eurostat has created an adequate instrument: the Centres and Networks of Excellence (ESSnet) projects for putting together expertise distributed throughout the ESS organisations in order to develop specific actions beneficial for the whole system.”

15. In the same document we find the first quotation of “our” ESSnet: under the title “Development and maintenance of shared IT tools for use in Member States” with the following 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.

16. 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.

17. 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.”

### B. ITDG 2008

18. The ITDG 2008 supported the launch of an ESSnet project on a common business architecture with the aim to create the enabling conditions for sharing software within the ESS; the document “ESSnet on a common reference architecture - preliminary ideas for discussion“ (you can find it at http://circa.europa.eu/Public/irc/dsis/itsteer/library?l=/directors_13-14/proposal_essnetdoc/_EN_1.0_&a=d) was the basis of our current work.

19. After the document discussion, Eurostat asked about countries interested in participating in ESSnet: Denmark, Italy, Latvia, Netherlands, Norway, Sweden, Switzerland expressed their willingness to join the ESSnet.

20. Eurostat agreed that ESSnet may assume the GSBPM model as a basis, besides CES 1999 document. It was recommended not creating new bodies, but aligning processes in MSIS and starting work as soon as possible to be able to progress before next MSIS meeting.

21. The need for consolidating of work on the software sharing was confirmed and consequently the importance of the work. The ESSnet shall be based on well defined principles and involve the business units. The ITDG agreed on the importance of software sharing as a way to optimize resources and to use tools based on sound methodologies. The approach to be followed should be pragmatic, based on designing a set of principles and rules for development and adaptation of tools to each environment.

### C. This year

22. In 2009 January in Luxembourg we had a preparatory meeting between interested countries and Eurostat for ESSnet: in the meeting, besides many administrative clarifications, we started to consider a layer architecture for introducing abstraction levels ensuring isolation between different layers and topics
In March 2009 METIS dedicated one session to GSBPM model (see

http://unece.org/stats/documents/2009.03.metis.htm ) in which the model was described and discussed. From Lisbon we received also some recommendation for our ESSnet:

- Common architecture outlines
- Adaptable to change
- Common integration rules
- Support for Open and Proprietary solutions
- Modular itself
- Integration of existing software
- First step to strategic goal of interoperability and easy interchange of solutions

## D. CORA!

23. In April 2009 we at least received MBGA and the work group started to work. Here we show a first draft of the Work-Packages structure and their goals:

WP1 - **Project Management**: *standard* project management WP

WP2 - **Requirements Collection and State of the Art**: definition of the overall set of CORA requirements through a questionnaire to be sent to NSIs and analysis of the State-of-Art (through collection of documents, classification and analysis of such documents in order to point out deficiencies and needs for extensions)

WP3 - **Design of the Statistical Data Lifecycle**: starting from previous proposals (notably GSBPM, but also Unece 99 and others), define a new proposal that addresses limitations of previous ones (generalisation, obsolescence, other dimensions)

WP4 - **Design of the Technical Architecture:** define layered architecture scheme from GSBPM down to IT Infrastructure level, describing levels/interfaces definition and standards for data and services integration.

WP5 - **Design of the Cooperation Architecture**: define the cooperation architecture in terms of licensing models for sharing applications, legal frameworks for shared development, business models for the support of shared applications, tools to be used .

WP6 - **Project Dissemination**: dissemination strategy of the project in terms of CORA web site, training courses on the products of the WPs, presentation of CORA results to meetings, organization of a “CORA ESSnet workshop”, presenting the conclusions of the Action

The work just started ...