

Distr.
GENERAL

WP.36
14 April 2010

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

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

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

Meeting on the Management of Statistical Information Systems (MSIS 2010)
(Daejeon, Republic of Korea, 26-29 April 2010)

Topic (ii): Software sharing and shared maintenance

Intermediate results from CORA ESSNet

Prepared by Carlo Vaccari, ISTAT, Italy

I. Introduction

1. Common Reference Architecture (CORA) is an ESSNet (European Statistical System NETWORK) funded by Eurostat to start the design of a technical and organizational architecture usable by Statistical Organizations. CORA was already discussed in Oslo 2009 MSIS meeting. CORA network is composed by seven European countries (Denmark, Italy (coordinator), Latvia, Netherlands, Norway, Sweden, Switzerland) and activities started in October 2009. Until now three meetings have been held and a survey was conducted. Final results will be presented in an open meeting-seminar in Luxembourg in October this year.

II. CORA work-packages

A. General matters

2. A common reference architecture brings several benefits to the NSIs. Firstly, it promotes a stronger *cooperation* among the different NSIs. Indeed, the different architectures are often an obstacle in this direction because they prevent from a real interoperability. Secondly, CORA improves the *sharing of solutions* among the different NSIs with considerable benefits in terms of cost reduction. Indeed, what typically happens is that each NSI develops its own software solution to address a problem that is present also in other NSIs. The reuse of this solution by a another institute is not always possible whereas technical architectures are different among each other. Thirdly, CORA stimulates a *quality improvement* of existing solutions for both business and technical layers of NSIs. Indeed, a shared architecture will be naturally designed according to the best state-of-the-art principles and paradigms.

3. The business architecture has the purpose of modelling statistical processes performed by NSIs as well as the underlying data life cycle. CORA starts from the GSBPM¹ and provides the necessary updates and extensions. The business architecture identifies a set of *logical blocks* composing the statistical processes. Such blocks have the purpose to be mapped into *technical blocks*, thus bridging the business and the technical dimensions of CORA. By identifying blocks as basic components of the CORA dimensions, we put the bases for a flexible architecture that can be easily adapted to changes that may occur. Specific *organizational rules* are defined for the logical blocks thus bridging the organizational and business dimensions of CORA.

4. Each layer consists of several technical blocks defined in terms of services delivered to the layer immediately above and of services requested from the layer immediately under. Services are defined in terms of communication interfaces between the layers. The definition of the technical blocks will be independent on existing software solutions, with the aim of judging whether these solutions are suited to their aims and of supporting the decision to reuse efforts already made by the statistical community, or to develop solutions that are more appropriate, and of helping the choice between equivalent solutions.

5. The technical blocks will specify the technical characteristics that implementation of such blocks should take into account. Moreover, technical blocks will consider open source solutions as a principal option for implementation. Whenever possible, proprietary solutions will be indicated and compared to the open source ones in terms of general features and appropriateness for specific usages.

6. The organizational architecture is defined in terms of several rules that must be respected when developing shared applications. Such rules include:

- (a) legal aspects;
- (b) licensing models for sharing applications, and
- (c) business models for the support of shared applications, with a specific attention to open source software development models.

7. The work of the ESSNet is divided into five work-packages: two WPs, namely WP3 and WP4, are dedicated to the implementations of the objectives related to the design of the architecture as a result of an integration of the three composing dimensions. WP2 is dedicated to the collection of previous work, as well as to a collection of requirements from different NSIs, not only the ones participating to the ESSNet. Finally, WP1 and WP5 are respectively devoted to the management and the dissemination activities.

B. Work-package 1: Project Management

5. This work-package considers all the duties of administrative management including the preparation of the meetings, managing project financial statements, monitoring and quality assurance of the outcomes of the project. Until now three meetings have been held:

- Kick-off meeting in Luxembourg, October 2009;
- WPs Leader Meeting in Amsterdam, January 2010;
- Intermediate Meeting in Rome, March 2010.

6. With respect to the deliverables, the Preliminary Report (as every deliverable) has been published on the CORA wiki, on the CORA web site and on the OSOR portal. As far as future work, it is programmed a Final Meeting to be done in Luxembourg during October. During this meeting, the CORA ESSNet results will be presented to Eurostat and to all others NSIs that will be invited to participate to this workshop. Always on October the Final Report will be ready.

C. Work-package 2: Requirements Collection and State of the Art

7. The objectives of this WP are the collection of requirements for the WP3 and WP4, and the production of a state of the art with respect to the architectures and the software products currently adopted by European NSIs. For this purpose, a questionnaire has been designed (deliverable 2.1) and distributed to 42 NSIs. During

¹ GSBPM: <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

the Intermediate Meeting in Rome, a response rate of about 79% (33 of 42 NSIs) has been reported from WP4 coordinator (Norway).

8. A first analysis of the compiled questionnaires outlined that, though several tools have been declared as “shared”, the actual usage of such tools by different NSIs did not emerge from responses. Moreover, the analysis outlined that the usage of an official statistical business process model was indicated by only four EU NSIs.

9. Future work of WP4 includes continuing the analysis of the received questionnaires together with the analysis of the documents received as part of the response to the questionnaire. Moreover, some quality checks on questionnaires results will be done, especially with respect to software tools.

10. Finally, the responses to the questionnaire will be summarised in a Set of requirements report that is the deliverable 2.2 (foreseen on May). Always on May will be available the deliverable 2.3, containing the documents classified and analysed to point out deficiencies and needs for extensions.

D. Work-package 3: Technical Architecture

11. The work done until now has led to the definition of a technical architecture based on a layered approach, in which lower layers offer services to upper ones. This approach has the advantage of providing clear “contracts” between the components at each layer that in this way have precise duties and rights. The technical architecture is based on a two dimensional approach.

12. The primary dimension, called the construction dimension, is determined by the way services make use of one another to deliver their respective products (data and metadata). Specifically, the layers are: (i) figures, a domain of interest documented by statistical products; (ii) time series, statistical series over time; (iii) statistic, integrated or simple statistical product for a given time; (iv) population, a population at a given time; (v) unit, a statistical unit at a given time; (vi) variable, a statistical variable at a given time; (vii) representation, a logical representation of the value of a variable.

13. The second dimension, called functional dimension, is formed by the categories of the GSBPM. Specifically, the layers are: (i) specify needs; (ii) design; (iii) build; (iv) collect; (v) process; (vi) analyse; (vii) disseminate; (viii) archive; (ix) evaluate.

14. One of CORA goals is to offer a model in which the GSBPM sub-processes of Level 2 can be described as services delivering a certain functionality within a certain layer.

15. As future work, beyond the refinement and the finalization of the CORA architecture proposal, a proof of concept activity has been added to WP3 activities because, being the CORA architecture a non-theoretical model, is not possible to test its correctness in a theoretical way. Hence, by developing a prototype we want to empirically test the validity and feasibility of the model.

16. Moreover, future works are: the definition of interfaces in order to guarantee the cooperation of the different layers; writing of a manual (deliverable 3.3) containing the instructions needed for a correct usage of the model; production of deliverable 3.4 containing both example mappings and proof-of-concept software prototypes

E. Work-package 4: Organizational Architecture

17. With respect to the first task of the WP, i.e. the definition of suitable licensing models of sharing applications, a thorough analysis of the available licensing models has been done, detailing characteristics, obligations of the licensees and correspondences between different licenses. Moreover, an analysis of the business models for training, support and development has been done. Examples of these models are: Barter

model, co-development, freeware, OSS development/support. Finally, a list of query and answers (Q&A) on legal aspects of the licensees has been realized and will be part of deliverable 4.1.

18. Next work includes the analysis of the technical and commercial requirements for tools that will be used by the CORA users; such an analysis has already started. This work needs WP2's results as input and is coordinated with WP3 work. The tools will be classified according to functional and non-functional categories.

19. Finally, as next work a list of available tools for each phase of a statistical process will be realized. After the publication of this list, its maintenance will be on charge of the MSIS - Sharing Advisory Board (SAB). As the previous one, this work needs WP2's results as input and is coordinated with WP3 work.

F. Work-package 5: Project Dissemination

20. With respect to the dissemination of CORA results at international meetings, the project has been firstly presented at the ITDG 2009 in Luxembourg, presenting mainly the objectives of the project and the organization of the work. Moreover, the final results of the CORA ESSNet will be presented at the final meeting that will be organized as an open workshop where all the NSIs are invited.

21. A cooperation with the OSOR (Open Source Observatory and Repository) portal started opening a project on the portal (<https://www.osor.eu/projects/cora>). Moreover, a CORA web site has been realized (<http://cora.forge.osor.eu>) always hosted by OSOR to reinforce the cooperation: in the web site all project deliverables will be made available for everybody.

22. Finally, a CORA wiki has been set-up, hosted by Istat (<http://www.wiki.istat.it/>); the wiki is accessible only to the participants of CORA, to the observers (Australia, Portugal and UK) and to Eurostat, being a working area containing draft documents,, provisional deliverables and areas in which partners can share ideas, documents, comments, working-on deliverables related to each work-package. To obtain the maximum of usability for CORA results, all deliverables will be released under Creative Commons license by-sa (<http://creativecommons.org/licenses/by-sa/3.0/>).

III. Future works

23. CORA final results will be presented in an open meeting in Luxembourg: to obtain the maximum of participation it was decided to organize the meeting the day before 2010 ITDG (October 18th), so that people attending to ITDG can minimize travel time and expenses.

24. CORA final results will be used also by MSIS Sharing Advisory Board: the software repository collected in the survey will be in the future managed by SAB in the UNECE wiki. Links will be established between the SAB repository and the OSOR project to optimize common efforts.

25. CORA project will design and demonstrate the feasibility of a common layered architecture, even through the development of some software prototype, but the defined architecture will need to be implemented in "real" software and in "real" systems. Eurostat will decide on a possible new ESSNet (perhaps CORE?) that will continue and complete CORA work.
