Distr.
GENERAL

WP.7
14 May 2012

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 2012)
(Washington, DC, 21-23 May 2012)

Topic (i): Architecture

Implementation of the ESS Joint Strategy [1]: "the development and promotion of the plug and play concept as an architecture model"

Supporting Paper

Prepared by Jean-Marc Museux, Nico Hilbert, Georges Pongas, Eurostat

I. Background/scope

1. The ESS Vision on the production of EU statistics [2] calls for industrialization of EU statistics production processes. This vision is common to the statistical industry as a whole. It has been identified as a key initiative by the High Level Group on Business Architecture for Statistics [3].

2. The ESSnet CORE (COmmon Reference Environment) [4], co-funded by Eurostat, has demonstrated that it is possible to define a production architecture which allows designing a statistical production system composed of several processes by assembling a set of independent services. The CORE architecture can be implemented in a platform independent and possibly distributed architecture model. It is sometimes referred to as "plug and play" architecture. This approach allows parameterisation of the various services fostering flexibility and reuse of software components in the definition of statistical processes. Services can be developed independently or cooperatively and shared among ESS partners. The concept is built on a common information model that allows communication between and automation of the tasks. The automation is based on a process registry which stores and makes available all production processes.

3. The need of an approach such as CORE is necessary for the following reasons:

(a) The rapid change in users' needs and the need to adapt our production methods to cope with new sources and types of information;
(b) The need to reduce development and maintenance cost;
(c) The need to reduce processing costs in the domains of software development and ad hoc human interventions;
(d) Whenever it is necessary to adapt a production environment to make it more agile for the development of new production lines, to allow integrating services/tools developed by partners;

(e) The improvement of quality of EU statistics through the even further harmonization of production methods and their systematic documentation.

4. The further development of the CORE architecture and the implementation in Eurostat and in the ESS has been identified by Eurostat as a strategic axis for investment in the five years to come. This is being discussed in the ESS and further development could start in 2013.

5. Although the technical and financial benefits of a CORE like approach are easy to defend, the adoption of the CORE model will imply a cultural change in the way we design our production and information systems. Given the rapid ICT developments and the need to ensure current production, a big bang approach is not feasible. These changes have to be prepared by a set of coherent activities in the different architectural layers of a Common Reference Architecture (CORA).

6. In Eurostat, few key reengineering projects (National Account Production System and Price Statistics Production) are planning to implement a modular approach based on SOA standard. Their alignment with CORE standards and their contribution to the further development of CORE standards are being considered.

7. This paper aims to describe initiative Eurostat intents to take to follow up on the initial CORA & CORE projects towards the development of the CORE model, its industrialization and deployment in the ESS.

II. CORE: Principal Results

8. The ESSnet CORE (for more detailed reference see [4]) started on January 2011 and ended on January 2012 with a total duration time of 13 months. It involved six NSIs as participants, specifically: Istat (Italian National Institute of Statistics) as coordinator, CBS (Statistics Netherlands), SSB (Statistics Norway), INE (Portugal Statistics), SCB (Statistics Sweden), and INSEE (France). Moreover, CORE had external observers from Slovenia and from Australia.

9. CORE ESSnet continued the work of the ESSnet called CORA (COmmon Reference Architecture), which finished in October 2010. CORA had delivered an architectural layered model together with proof-of-concept software prototypes.

10. Starting from CORA results, CORE defined a complete information model that is a communication protocol for the exchange of information between a CORE service and its environment. More specifically, the protocol describes the information elements a service receives in order to configure it, subsequently the elements a service receives as input upon execution, and finally the elements a service offers as output after execution. In this way, the interface between two services, or between a service and a service execution mechanism (i.e. a run-time execution engine), is established. The information model is an abstract model composed of classes and relations that are meant to be stable in time. It is meant to support the definition of a more concrete model: the interface model, which is expected to vary over time.

11. The design of the CORE environment is another significant output of the project. The design firstly defined the interface model, and secondly the logical architecture needed to run processes according to such a model. More specifically, the CORE environment permits: (i) the definition of statistical processes in terms of abstract services which are GSBPM-labelled; (ii) the design of data and metadata defined as input/output of services, (iii) the execution of designed processes by invoking existing (wrapped) tools. The strength of this environment is the ease of use by statisticians: CORE actually developed interfaces (GUIs) to achieve this objective.

12. The current CORE software implementation consists in a prototype showing that it is feasible to implement the CORE architecture design. More specifically, the general principles proven by the prototype can be summarised as follows: (i) it shows a real implementation of industrialized (standardized and automated) statistical processes; (ii) it enables the reuse of IT tools developed on different platforms and by different NSIs; (iii) it is a first example of a GSBPM-aware services implementation; (iv) it provides a unique common data

model enabling integration of heterogeneous data exchanged between services; (v) it supports openness to evolving statistical information models (e.g. GSIM).

III. Towards the development and adoption of CORE model in the ESS: first list of issues

A. Maturation of the CORE conceptual architecture

13. The bridging between the CORE data model framework and GSIM, which is currently under development, (in particular its process component) should be further pursued as a step towards internationalization.

B. Maturation of the CORE technical architecture

14. The further development of the CORE model and its industrialization will have to go through the development of real size prototypes allowing the full evaluation (cost benefit, performance) of the concepts and the strengthening of business cases before developing strategic plans for its deployment.

15. Of particular importance are the following issues:

(a) the evolution of the data model and its link to SDMX, DDI or possibly the matrix model used in Banca of Italia [4],
(b) the definition of the metadatabase having the capability to describe the inputs, outputs, constraints and calling sequences,
(c) the choice and integration of a process monitor.

16. The prototype development should be guided by different principles like the variety of underlying infrastructure and should cover local and distributed environments.

17. What is more, the selection of a statistical process should be supported by a strong business case:
For the legacy system: the candidate process is homogeneous enough, time consuming, and for which generic software already exists using web services (e.g. editing, tabulation). It would be possible to cover a part of the full process.
For a new system: the chosen process should be of subcritical importance but should encompass most of the anticipated innovation in the statistical process (integration of new sources, flexible dissemination …)

C. Maturation of CORE organisational architecture

18. As previously said, the full deployment of the CORE architecture would be facilitated if NSIs adopted a more functional organisation. The ESS Sponsorship on Standardisation is reviewing the ESS business process to identify where the standardisation would yield the highest benefit and possibly identify where a functional approach could be beneficial. Preliminary Eurostat analysis points towards process steps requiring significant interaction between stakeholders such as for the use of administrative data, EU statistics validation, data collection/reception, common register service etc..

19. The international collaboration for software sharing initiated through the CORA project should be pursued and extended to identify and to document the CORE compliant services.

20. In the context of the ESS, the business model and the funding mechanism for the development and maintenance of services should be developed.

21. For the CORE technical governance, an international steering group could be set up for reviewing and analysing the CORE development.

D. Preparing for CORE deployment

22. In parallel, efforts will be needed for defining, promoting and adopting a set of CORE principles for IT service development (platform independence, service orientation, interoperability with identified data/metadata standards on top of traditional principle like robustness, scalability, …) for the IS developments in the Statistics Community.

F. International collaboration

23. Despite the ESS benefits from the deployment of CORE model and architecture, it is fully understood that the CORE developments should benefit the whole Statistics Community. The compliance with higher level standards like GSIM and GSBPM and current technical standards like SMDX is a must. On the other hand, the identification and the adoption of the subset of standards underlying the CORE model is necessary to ensure international interoperability and reuse of CORE components. The involvement of external partners to the CORE technical development steering group is also a minimum requirement.

IV. Next Steps

24. Eurostat has proposed to convene the former CORE partners for a brainstorming and the establishment of a roadmap. This roadmap is necessary to mobilise the required ESS funding for the next five years. The roadmap should be discussed at the HLG-BAS workshop in Geneva in November 2012. The relevance of the concept for the ESS will be further worked out with the relevant ESS working groups (SISAI in June 2012, ITDG in November 2012).

V. References

[1] http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/about_eurostat/documents/ESSC20100506ENvisionfinal.pdf

[2] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the production method of EU statistics: a vision for the next decade http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF

[3] High Level Group on Business Architecture for Statistics: http://www1.unece.org/stat/platform/display/hlgbas

[4] CORE (COmmon Reference Environment) project final report http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/24_Italy_CORE.pdf

[5] http://www.bancaditalia.it/statistiche/quadro_norma_metodo/modell_SIS/StatisticalDataAndConceptsRepresentation.pdf