Distr.
GENERAL

WP.24
9 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 (iv): Collaboration

CORE (COmmon Reference Environment) project final report

Supporting Paper

Prepared by Carlo Vaccari, Istat - University of Camerino

I. Introduction

1. The ESSnet CORE (COmmon Reference Environment) 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[1] (Italian National Institute of Statistics) as coordinator, CBS[2] (Statistics Netherlands), SSB[3] (Statistics Norway), INE[4] (Portugal Statistics), SCB[5] (Statistics Sweden), and INSEE[6] (France). Also two universities were involved as third-parties: University of Athens[7] (Greece) and University of Camerino[8] (Italy). Moreover, CORE had the following external observers: Mr. Tomaz Speh (Slovenia) and Mr. Alistair Hamilton (Australia).
CORE ESSnet continued the work of a previous ESSnet called CORA[9] (COmmon Reference Architecture), which finished in October 2010, and delivered an architectural model together with proof-of-concept software prototypes.

---

1 http://www.istat.it/en/
2 http://www.cbs.nl/en-GB/menu/home/default.htm?Languageswitch=on
3 http://www.ssb.no/english/
4 http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_main&xlang=en
5 http://www.scb.se/default____2154.aspx
6 http://www.insee.fr/en/default.asp
7 http://en.uoa.gr/
8 http://www.unicam.it/international/welcome/index.asp
9 CORA deliverables can be found at http://www.essnet-portal.eu/finished-projects/cora-finished

2. 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 in time. This two-step approach has been chosen because it allows reduction of the maintenance burden of a model that will inevitably be subject to change as it is confronted with the concrete reality of managing statistical processes.

3. The *design of the CORE environment* is a further significant output of the project. The design included first the definition of the interface model and then the definition of a logical architecture implementing the definition and running of processes that conform to such a model. More specifically, CORE environment permits: (i) the definition of statistical processes in terms of abstract services GSBPM-labelled; (ii) the design of data and metadata defined as services inputs/outputs, (iii) the execution of designed processes by invoking existing (wrapped) tools. The strength of this environment is the ease of use by statisticians: we have projected and developed interfaces (GUIs) that permit the full achievement of this objective. The *CORE implementation* is a prototype that shows the feasibility of implementation of CORE architecture design. More specifically, the general principles proved by such a prototype can be summarised as: (i) it shows a real implementation of industrialized (standardized and automated) statistical processes; (ii) it enables a reuse of IT tools possibly 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).
During this project, we have also evaluated the costs related to the integration aspects in order to provide other NSIs with an indication of the costs they will have to bear in order to adopt the CORE environment into their institutes. Finally, we have studied mapping issues between SDMX and the CORE model; the results of this analysis are published in a dedicated deliverable (deliverable 3.3 “Lessons learned on the use of SDMX within data processing”).

II. CORE work-packages

A. Work-packages introduction

4. The work of the ESSnet has been structured into the following work packages:
- WP1: Project management;
- WP2: Design of the information model according to GSBPM[10] and alignment with NSI's information models;
- WP3: Generic interface design for interconnecting GSBPM sub-processes;
- WP4: Research workflow solutions for process management;
- WP5: Implementation library for generic interface and production chain for .NET;
- WP6: Implementation library for generic interface and production chain for Java;
- WP7: Project dissemination and integration cost reporting.

In Figure 1 the participation to each WP by each partner is shown.

---

10 Last version of GSBPM can be retrieved at http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model

	Istat	CBS	SSB	INE	SCB	INSEE
WP1	**C**	P	P	P	P	P
WP2	P	P	P	P	P	P
WP3	**C**	P	P			P
WP4		P	P	P	P	**C**
WP5		**C**				
WP6	**C**			P		
WP7	P	P	**C**	P	P	P

Figure 1: Participation to WPs (C=coordination, P=participation)

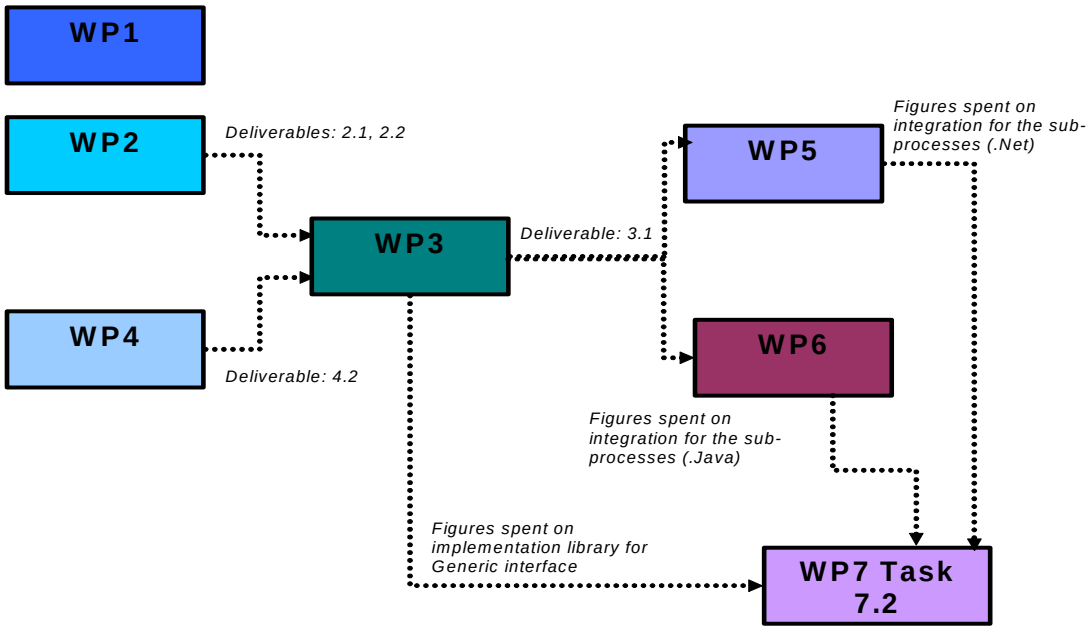

Figure 2. WPs dependencies

Figure 2 shows the existing dependencies between WPs. In more details, the results of WP2 and WP4, specifically deliverables: 2.1, 2.2 and 4.2, are preparatory for WP3 which is in turn preparatory (deliverable 3.1) for WP5 and WP6. Finally, WP3 (activities spent on implementation library for generic interface), WP5 (activities spent on integration for the sub processes in .Net) and WP6 (activities spent on integration for the sub processes in Java) are preparatory for the second task of WP7 that is the task devoted to the integration cost reporting. The WP1 is devoted to the management and the first task of WP7 is devoted to the dissemination activities.

In the following sections, the work done for each WP is detailed.

B. WP1 Project Management

5. WP1, coordinated by Istat, started on month M1 and ended on month M13. This WP has been organized into two main tasks, the former including organizational and administrative activities, and the latter including quality control on project deliverables. Both tasks were carried out by means of emails and web conferences, in addition to face-to-face meetings planned as the official project meetings.
With respect to the project meetings, as planned at the proposal stage, the following meetings have been organized:

(a) Kick-off meeting - Luxembourg, January 2011;
(b) Lisbon meeting – Lisbon March 2011;
(c) Intermediate meeting – Paris June 2011;
(d) Rome meeting – Rome October 2011;
(e) Final meeting – Luxembourg, January 2012.

During the final meeting the CORE ESSnet results have been presented to Eurostat and to other NSI representatives, including Austria, Estonia, Luxembourg and Slovenia. This WP has produced four deliverables: 1.1 Preliminary Report, 1.2 Intermediate Report, 1.3 Final Report and 1.4 Cost Report.

C. WP2: Design of the information model according to GSBPM and alignment with NSI's information models

6. WP2, coordinated by Statistics Netherlands, had the main objectives to design a generic information model for the statistical process and to develop a method to map it onto the business model of any NSI.
To this purpose a depth requirements analysis has been done: WP3, WP5 and WP6 were seen as internal stakeholders while ESSnet observers were seen as external stakeholders. The needs of these NSIs have been analysed and features have been specified to meet their requirements. Deliverable 2.1 Requirements, contains the results of this analysis.
In accordance with these requirements, a generic statistical information model was designed and published (deliverable 2.2 Generic statistical information model).
A generic mapping method between the information model and business models used by NSIs cannot be defined as a direct relationship. This relationship can be established via the GSBPM. A reference to GSBPM is part of the information model. Therefore, the required mapping has to go through a mapping between the GSBPM and the business models of individual NSIs. This mapping was delivered in the form of several case studies in chapter 3 of deliverable 3.2 of the CORA project. This approach is presented in deliverable 2.3 Mapping methodology.
Given that a number of participants at conferences and workshops in which CORE was presented have asked for a document explaining how to use the results of the project, deliverable 2.4 has been redefined as “How to use CORE”.

D. WP3: Generic interface design for interconnecting GSBPM sub-processes

7. The main objective of this WP, coordinated by Italy, was the design of a software environment able to support the information model designed in WP2. Starting from the WP2’s results, we realized a complete technical definition of the environment and gave a detailed description of how data are exchanged. The results of this WP have been described in deliverable 3.2 “Technical Environment Specification vs2”. A previous deliverable (3.1 Technical Environment Specification vs1) containing the first results of this work was delivered

in the middle of the project. Moreover, this WP had the objective of studying the way of mapping SDMX[11] and the CORE model; the results of this analysis are published in deliverable 3.3 "Lessons learned on the use of SDMX within data processing".

8. The overall design of the CORE environment includes the following main logical components:

(a) A Graphical User Interface (GUI) component providing a set of components for: *(i)* process specification, according to a defined process modeling language, *(ii)* service definition and *(iii)* specification of data to be exchanged among services.
(b) A set of components, called Integration APIs, wrapping existing (possibly legacy) software tools owned by NSIs. The specific purpose of Integration APIs is the conversion from/to tool-specific data to/from CORE data.
(c) A Process Engine in charge of the execution of processes specified in terms of defined services.
(d) A Definition Repository storing: *(i)* process schemata, that report the choices made in the process specification phase, *(ii)* service specifications, in terms of where services to be called are, where input/output of such services should be taken from, etc., *(iii)* data models, consisting of the defined schemata for data exchanges within CORE environment and *(iv)* data, comprising all the data that must be passed to CORE services as input/output, as well as housekeeping information useful for the process management.
(e) Services and their runtime, permitting the execution of the overall CORE process "composed" by the available services.

Moreover, a concrete process scenario was designed to be used as an empirical test-bed during the whole implementation cycle of the CORE environment. Such a scenario also included a conversion from CORE data to SDMX data to analyze issues related to the integration of the two models.

E. WP4: Research workflow solutions for process management

9. WP4 was coordinated by INSEE. As first milestone, we produced an evaluation grid to compare the different tools that are available to manage the interactions between services designed. This evaluation grid contains different dimensions such as Integration (versioning, user management, communication), execution (monitoring, traceability), general issues (use of standards, automatic and manual steps) and modelling (GUI, model translation) useful to evaluate and compare the different tools. Deliverable 4.1 Evaluation grid used to evaluate the different tools contains the results of this milestone.
As second milestone, we designed a process scenario to evaluate tools so that we found efficient tools to support the activities of interactions between services designed. We have studied both open source and proprietary solutions in order to have an assessment of the state of the art. Specifically, we analysed three different solutions: Activiti, Bonita (both open source solutions) and ActiveVOS (a commercial product). The results of this work and the choices made are contained in deliverable 4.2 "Recommendation on process management tools".

10. The goal of the WP4 was to search for efficient tools to support the modeling, execution and monitoring of statistical processes composed of CORE services. These tools are usually referred to as BPM (Business Process Management) tools. The first task of this work package was to achieve a common understanding of the BPM software panorama and of CORE needs in this field. This was realized at the Lisbon meeting, where a requirement list for the tools was adopted. Then INSEE created a detailed evaluation grid according to those requirements. A second task was to elaborate a short process scenario to be used for evaluating the tools. The scenario represented the calculation of an aggregated statistic in three steps using different technologies and protocols: a web service call, a manual validation (with a possibility of rejection looping back to the service

[11] http://sdmx.org/

call) and a language-specific program execution (Java archive or .Net application). This scenario was approved during Paris meeting, and mock web services and programs implementing the first and last steps were written shortly after. The evaluation phase began after the Paris meeting, where a first list of three candidate products was validated (for various reasons, one of them was later replaced by an alternative solution). Two of the candidates (Activiti and Bonita[12]) were prominent open-source initiatives and the third one was a commercial product (ActiveVOS[13]), included in the list as a point of reference and because it had been chosen by the Australian Bureau of Statistics. Due to the limited amount of resources, we focused on Java software. The evaluation procedure consisted of several steps: reading all the documentation, installing the software, modeling the scenario, executing it (while acting as the business user for the human step) and monitoring the execution. During the whole process, the evaluation grid was filled in and a textual report was completed. The final report was presented at the Rome meeting. The conclusions of the study were validated by all participants. They showed that, although ActiveVOS was as to be expected in first position for the functional coverage, product quality and general user experience, Bonita was a very strong contender: it is easy to install and use, very versatile and open. Activiti, although very promising, was judged too young and ranked last.

F. WP5: Implementation library for generic interface and production chain for .NET and WP6: Implementation library for generic interface and production chain for Java

11. WP5 were coordinated by CBS and WP6 were coordinated by Istat. These WPs both have the aim of providing a working solution that permits the automatic execution of selected GSBPM sub-processes starting from WP3 and WP4 results. WP5 provided an implementation in .NET while WP6 provided an implementation in Java that are respectively proprietary and open languages.

12. The proof-of-concept of the CORE runtime environment was first (as a draft) implemented as a Bonita solution running on the Statistics Netherlands platform and then translated from WP6 into a Microsoft.NET application, developing following components:

(a) An integration API for managing CSV-to/from-CORE conversion, implemented in C#.
(b) A “basic” process engine in charge of the execution of sequential processes, implemented in C#.
(c) The definition repository, implemented in C#/SQL Server.
(d) Services implementing the selected real scenario implemented in C#.

Development of a design environment was out of scope; the designs were produced by the design environment from WP6.
The objective of WP6 was to provide an implementation showing the feasibility of design of the CORE architecture.

13. The proof-of-concept of CORE environment was implemented by open source software solutions, mainly Java and MySQL. More specifically:

(a) GUIs for process, service and data specification were implemented in Java. An open source framework for process design was also evaluated, namely Oryx[14]. An ad-hoc extension of Oryx tailed to CORE was also implemented;
(b) An integration API for managing CSV-to/from-CORE conversion was implemented in Java;
(c) A “basic” process engine in charge of the execution of sequential processes was implemented in Java;
(d) The definition repository was implemented in Java;
(e) Services implementing the selected real scenario were implemented in Java.

---

[12] http://activiti.org/ and http://www.bonitasoft.com/
[13] http://www.activevos.com/
[14] http://bpt.hpi.uni-potsdam.de/Oryx/WebHome

G. WP7: Project dissemination and cost reporting

14. The main objectives of this WP, coordinated by SSB, were the dissemination activity of the CORE results and the integration cost reporting. With respect to the dissemination of the results we have realized:

(a) CORE website hosted by Eurostat: www.essnet-portal.eu/project-information/core
(b) CORE wiki hosted by Istat using the tool DokuWiki: https://www.wiki.istat.it/doku.php/start?idx=essnet_cra.

Moreover, the CORE results have been presented at the following international meetings:

(c) New Techniques and Technologies for Statistics held in Brussels in February 2011 (NTTS 2011[15]);
(d) Joint UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems MSIS held in Luxembourg in May 2011[16];
(e) Statistical Information Systems Architecture and Integration (SISAI[17]) Luxembourg June 2011. At this meeting have been presented not only CORE results but also the results of joint work between CORE and SDMX ESSnet;
(f) Methodologies, Technologies and Tools enabling e-Government (MeTTeG[18]) held in Camerino (Italy) July 2011;
(g) Meeting organized by Senior Statistician of the Statistical Service of Greece, Crete July 2011;
(h) Working group for "Technology innovation" by the General Secretariat of Research & Technology, Greece, September 2011;
(i) Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS) Geneva October 2011[19]:
(j) ITDG Luxembourg October 2011[20]
(k) ESSnet workshop 2011 Cologne, Germany October 2011[21]

Finally, this WP has collected, assured quality and reported the amount of time spent on integration for the sub-processes in WP5 and WP6.

[15] http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/NTTS_2011
[16] http://www.unece.org/stats/documents/2011.05.msis.html
[17] http://www1.unece.org/stat/platform/display/msis/SISAI
[18] http://conferences.cs.unicam.it/metteg11/
[19] http://www.unece.org/stats/documents/2011.10.metis.html
[20] http://circa.europa.eu/Public/irc/dsis/itsteer/library?l=/directors_18-19&vm=detailed&sb=Title
[21] http://www.essnet-portal.eu/meetings/essnet-workshops/koln-2011