Modernising European business statistics' infrastructure[1]

European Establishment Statistics Workshop

7-9 September 2009, Stockholm, Sweden

Authors

Axel Behrens, Marc Chovino, Claudia Junker, **Jean-Marc Museux**, Jan Planovsky and Hartmut Schroer – **Eurostat**, the Statistical Office of the European Communities

Contact: Jean-Marc.Museux@ec.europa.eu

Abstract

The paper reviews some of the initiatives coordinated by Eurostat that aim at modernising the infrastructure of business statistics in the European Statistical System (ESS). Initiatives like the MEETS programme, the Data Life Cycle project and initiatives related to ESS metadata, and standardisation of methods and tools are presented in perspective with the new business model proposed by the Commission aiming at integrating and rationalising ESS statistics production.

Introduction

On 10 August 2009, the European Commission adopted the Communication on the production method of EU statistics: "a vision for the next decade". The vision will serve as basis for developing ESS statistical production processes and will thus have an impact on the development of processes for enterprise statistics.

The communication is based on the following statement: "the way of producing statistics in numerous parallel processes, country by country and domain per domain, following the traditional stovepipe model is no longer fully adapted to the changing environment". The driving forces for change are the globalisation of the economy, the development of Information Communication Technology, the need to produce more statistics and at the same time to reduce costs and burden and thus to improve efficiency by avoiding duplication of work and exploiting synergies.

The communication puts forward a more integrated approach of production of statistics drawing on a common data warehouse, the harmonisation of procedures through the sharing of IT tools, the identification of best methods and a shared dissemination.

The communication identifies three enablers for development of this new vision:

Firstly, a new governance system that allows for more integration: the European Statistical System Committee, the new body created by the new Regulation on ESS statistics to reinforce collaboration between the ESS partners, has the opportunity to streamline the regulatory environment for producing statistics by fostering their integration through the adoption of a common framework for developing statistics.

---

[1] Revised version 31.08.09

Secondly, the methodological developments and the definition of best practices will foster the harmonisation of processes and statistical output and create opportunities for standardisation.

Thirdly, new information technologies open new modes of communication with users (web 2.0), new ways to collect statistics and to share and maintain tools (open source development, service oriented architecture).

This paper aims to explore the impact of the new vision on the production of enterprise statistics, to review early developments that contribute to the setting up of the new integrated production process and to identify main challenges.

From an "augmented" stovepipe business model to a fully integrated business model for business statistics

Our description of production of business statistics is based on the 'Generic Statistical Business Process Model" (GSBPM)[2] developed jointly by UNECE, OECD and Eurostat. This model describes all steps in the process of statistical production from identifying user needs, data collection to the evaluation of the process.

In the context of ESS statistics, the process involves two types of actors: NSIs and Eurostat.

In the current model, illustrated in Figure 1, Member States are partners providing data to Eurostat which ensures the dissemination of EU statistics. Both at NSI and Eurostat level the production processes are broken down by the process steps described in the GSBPM. The ordering of the phases has to be adapted to the fact that ESS statistics user needs are defined at EU level. The different steps in natural order of appearance are

1) **User needs** identified at European level by **Eurostat** through requests from Commission DGs, the European Council, etc and through hearings with users in the Commission.

2) Eurostat, in collaboration with Working Groups (experts from the MS), integrates these needs and '**develop and design'** or adapt instrument. Minimum methodological requirements and the required variables and breakdowns and associated harmonised definitions are defined.

3) The respective statistical **Regulations** including the main methodological aspects and a few methodological guidance are usually adopted by comitology procedure. An overriding principle of the Regulatory process is subsidiarity. In enterprise statistics the regulatory context is mostly 'output-focused'. MS are obliged to respect the required output but the production of the output is done according to their national practices.

4) **NSIs** integrate EU and national needs to **develop and design** and then **build** or **adapt** their national instruments to **collect data**, to process and to analyse them. The analysed data are **transmitted** to Eurostat as the contribution to EU statistics.

---

[2] http://www.unece.org/stats/documents/ece/ces/ge.40/2008/wp17.e.pdf

5) In the meantime, **Eurostat** builds a specific infrastructure to **receive, validate and ex post analyse the data and finally disseminate** them. Secondary uses of the data at national level and EU level (access of micro data for research …) are not described here.

In the current model there can be as many processes and instruments as statistical domains and MSs. It goes without saying that this model, that seeks specialisation and reflects the natural organisation of statistical offices, generates also extensive use of resources (in development and maintenance). It is advocated that this model can be changed into a more integrated one finding a good balance between costs and quality.

The new model can be built on a few cornerstones, which would all foster integration of processes:

1) An **integrated Business Register** for enterprises (covering also multinational enterprise groups) used by the different actors of the statistical process.

2) A common **data warehouse** organised around consistent statistical units, gathering information across domains, where data from different sources (surveys and administrative) are integrated and linked at micro/meso level and made consistent.

3) The definition of **standard processes** based on **current best methods** implemented in **shared tools**. The level of standardisation may range from input harmonisation when relevant to output harmonisation based on common validation rules. The process should consider the development of smarter ways to collect business data making more efficient use of data within the ESS (**data integration**) and in the economy (**administrative data**). It will require more efficient tools for data and knowledge extraction, transmission and processing.

4) The setting up of a **harmonised system of metadata** across the ESS and across the domains covering structural metadata, such as dimension names, codes, hierarchies to facilitate data exchange. Reference metadata should also be harmonised by setting standards to improve quality. Ultimately, process metadata (such as validation rules) should also be harmonised to ensure full traceability and comparability of data, regardless of their origin.

5) A **Service Oriented IT Architecture**, defining a protocol for decentralised development of different services. Service could communicate through Web service technology. Prior to development metadata standard (XML scheme) and functional analysis of production processes should be established and documented. The Service Oriented IT Architecture allows for standardisation if the type of service is monitored and limited at ESS level. MS and Eurostat, can tailor their process to its their actual needs by using a series of shared and standardised components.

The new business model can be illustrated by Figure 2 where MS and Eurostat contribute to selected steps of the global process.

This long term vision can only be implemented step by step. In the following section, we review the early steps already taken, or in the process of being launched, which contribute to

the setting up of the new infrastructure. Particular importance is given to the MEETS programme (see background paper prepared for the panel session for an outline of the programme) which entails the largest potential of development for production of ESS business statistics in the coming years.

Review of on going projects contributing to new architecture for producing business statistics

Shared IT architecture and common tools

Over the years, in Eurostat, the initially homogeneous statistical production environment on central computers has evolved into a heterogeneous range of production systems in which practically every statistical domain (STS, SBS, Trade Statistics …) has its own specific IT system to collect, validate, analyse and process statistical data.

The CVD (**C**ycle de **V**ie des **D**onnées or Data Value Chain) project aims to provide a coherent set of concepts, metadata structures and IT tools to be applied in all statistical domains. Such a project creates conditions for having significant benefits, such as economies of scale in developing IT tools, improved functionality and a basis for key corporate objectives such as quality management and enhanced competence management.

In this context, several building blocks have been developed covering some of the production steps and available through the corresponding process handler. Currently Eurostat is developing building blocks for 1) data editing and derivation 2) statistical disclosure control 3) outlier detection 4) economic indices 5) user support 6) seasonal adjustment.

Eurostat next objective is to create conditions to extend the concepts at the ESS level. For this purpose, Eurostat has launched an ESSnet[3] project called "**common reference architecture for the ESS**"; it will study how to bind existing business and technical architecture to a reference architecture. While National Statistical Institutes mostly have the same or very similar statistical tasks to perform, NSIs have architectural differences, both to the business and technical side. The project will start in 2009 and it will run for 10 months with the following objectives:

a) describe a generic Business Process Model which would pave the way for harmonisation;

b) design Information Systems Architecture describing the logical architecture for data and metadata and applications at logical level;

c) identify a possible technical architecture, technical platforms and use of standards. The architecture proposed should allow for cooperative development.

The project is also expected to produce a more unified/integrated description of the processes between Member States and Eurostat to produce EU aggregates. The first contributions of Eurostat to the shared infrastructure are likely to be based on already existing building blocks adapting them to the ESS environment. In particular, there are projects to propose a web service based on the Eurostat Editing Building Block. It will allow MS to check their data

---

[3] ESSnet are ESS collaboration networks which consist of projects carried out by a team of institutions aiming at developing results which can be used by the whole ESS community. ESSnet projects are co-financed by the Commission and participating institutions.

against or better to integrate Eurostat rules in their process before sending the data through the single entry point, eDAMIS (the standard tool for secure and monitored transmission of data from providers to Eurostat).

Harmonisation, streamlining and common metadata

One of the objectives of MEETS programme is to streamline the framework of business - related statistics. An integrated system in the Member States requires coherent legislation over the domains at Community level. The focus will be put on the consistency between the statistical units in different domains and the common definition of target populations. At the same time the framework should allow for the use of administrative data; this might require to allow deviations from the target definition (concept and reference period).

The actions undertaken here should be integrated, prolonged and amplified by the development of a standard metadata. In all domains of business statistics the standard metadata can be used to identify, formally describe or retrieve statistical data; it consists of dimension names, variable names, dictionaries, dataset technical descriptions, dataset locations, keywords for finding data etc. For example, structural metadata include the titles of the variables and dimensions of statistical datasets, the statistical units, the code lists (e.g. for territorial coding), data formats, value ranges, time dimensions, value ranges of flags, classifications used, etc.

At Eurostat level, work is to create a single environment to store and manage harmonised structural and work on reference metadata is on-going. A reference metadata standard has been created at ESS level, building on the SDMX Information Model (part of the SDMX standard); it is designed to describe aggregate statistics as transmitted by MS to Eurostat. Standards for structural metadata are also under development through the harmonisation of standard code lists and, within SDMX sponsoring organisations, the harmonisation of the concepts contained in a number of code lists used at international level. As a result more and more harmonised code lists are produced and released by SDMX and Eurostat.

Business register and data warehouse

The MEETS programme aims to make full use of all data available in the statistical system. Data warehouse approaches will be investigated and developed from 2010 for 4 years aiming to develop fully integrated data sets for business statistics at micro level. Projects will also aim to develop a common methodology and identify best practices and to test the feasibility of the implementation for business statistics in MS and across Europe.

The development of the methodology of business registers on multinational enterprise groups is underway; this development in brought under the framework of MEETS since 2009. It will be implemented in the EuroGroups Register, the ESS common register of enterprise groups. A workflow between Eurostat and the National Statistical Institutes for the exchange of information will be defined and provisions for data quality management and maintenance will be set up. The development of national infrastructure will be funded. The profiling of the Large and Complex Multinational Enterprise groups will tested from 2010.

All the know-how and infrastructure created around the EuroGroups Register project can be seen as the basis for an Register of EU enterprises adapted to the renewed framework for EU business statistics.

Smarter way to collect data

A methodological ESSnet should be launched in 2010. It will identify best practices, develop common methodologies and produce ESS guidelines to support the production of business statistics. Its aim is to reduce respondent burden and to foster efficiency and integration of processes. The expected output is a common set of methods for the coordination of samples and surveys, the optimal design of data collection, the integration of different sources and model-based estimation methods. This action has to be seen in a more global framework supported by thus the ESS Group of Directors of Methodology (DIME) which aims at identifying and developing best methods through cooperative developments in order to facilitate the integration and standardisation of processes. Through searching scientific consensus on new methodological domains, they also pave the way for developing new common tools.

Implementation work on data integration will be carried out specific projects, for instance linking of micro-data on international sourcing and structural business statistics, linking of ICT statistics and business statistics and linking of business and external trade statistics. This work can draw on the results of a first ESSnet project on Integration of Survey and Administrative Data (ISAD) completed in 2008. The project has built a common methodological knowledge base for data integration (micro data linking and statistical matching) from different sources in the ESS. It will be continued in 2009 – 2010 putting more emphasis on micro data integration.

The use of administrative data will also be fostered by projects launched in 2009 on the better use of administrative and accounting data for business statistics; it seeks to develop recommendations and to analyse best practices when investigating the usefulness of available administrative data for business statistics by initial quality checks of the administrative data, and by combining sources for producing business statistics.

Actions in Member States related to the creation or the development of systems that facilitate the data transfer from enterprises to NSIs will be supported. They will aim at the creation of national statistical taxonomies based on accounting standards as well as the development of tools for extracting statistical data from companies' accounts. Actions can draw on The XBRL Pilot Project run by Eurostat together with a group of Member States in 2006-2007 that tentatively concluded it was possible to create a European statistical XBRL taxonomy incorporating the SBS and STS variables.

Conclusions

The implementation of a new business architecture for ESS business statistics production is a long process. Benefiting from the impulse of the MEETS programme, some actions in this direction are already on-going. The integration of ESS processes and the different on-going actions require the setting-up of 3 corner stones (listed by order of priority): 1) a high level description of statistical production processes specifying input and output for each step and current best methods, 2) a common metadata reference for the business statistics process, 3) the definition of a generic ESS IT-architecture on which future development can draw. The experience so far highlights the need to base methodological and infrastructure developments on cooperative network, which acts for the benefit of the whole ESS.
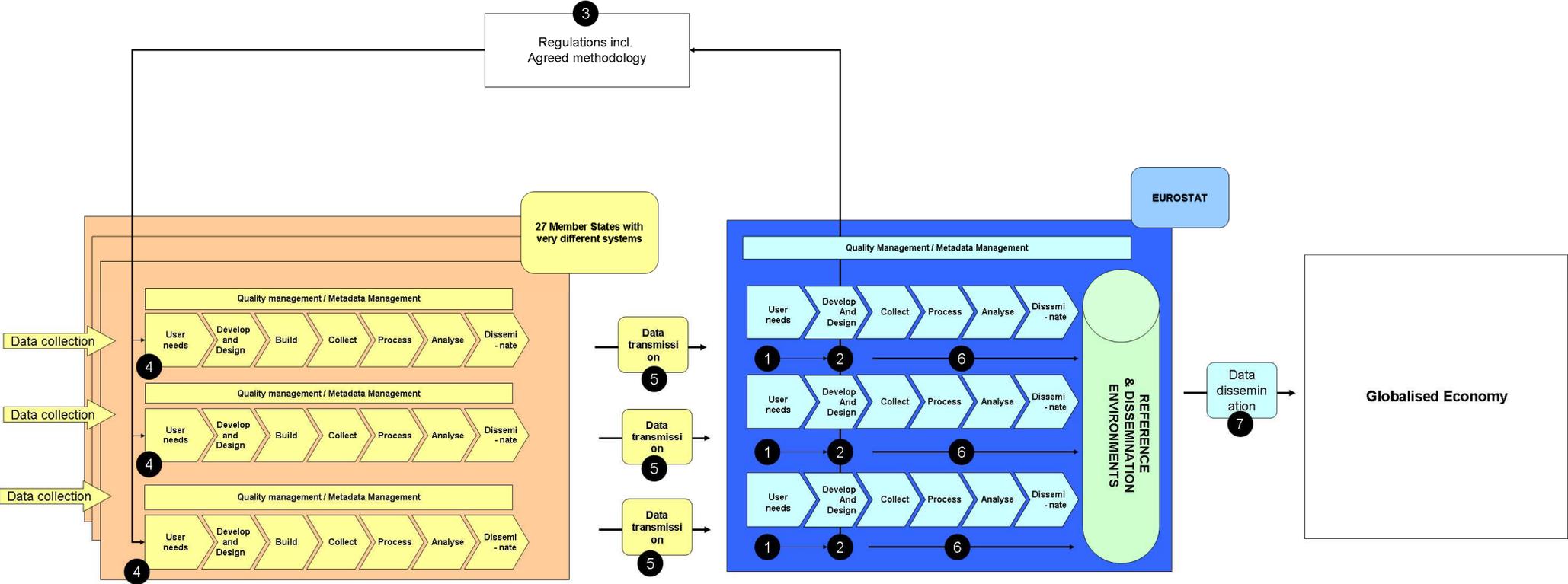
FIGURES

Figure 1 Stove pipe model

Figure 2 Integrated model