Meeting of the Group of Experts on Business Registers
organized jointly by UNECE, Eurostat and OECD

Brussels, 21-23 September 2015

Jean-Marc MUSEUX, Ioannis XIROUCHAKIS, Gergely KOEVESD
EUROSTAT
Session No. 4

The role of statistical business registers in the modernization of the statistical production and services – GSBPM, GSIM, data warehouse, use of new data sources, including big data

EuroGroups Register contributions to ESS Enterprise Architecture

Resume

The paper analyses from an EA architecture perspective the Euro Groups Register (EGR) functions and components in the context of the modernisation of the ESS system of Business Registers (ESBRs) and the ESS Vision 2020 implementation. After the presentation of the scope of the ESBRs project, the functions and applications constituting the EGR, the structure of the ESS EA and the forthcoming Business Service Catalogue, the paper highlights the EGR potential contributions to the modernisation of the ESS and the ESBRs using the ESS EA framework. The paper concludes discussing the expected benefits of such an architectural approach.

1. INTRODUCTION

1.1. The ESBRs project in the ESS Vision 2020

The goal of the ESS.VIP.ESBRs project (or simply ESBRs project) is to support the production of higher quality business statistics through increasing the consistency and coherence of statistical business registers (SBRs) across the European Statistical System (ESS). The main objective is to improve the interoperability of the SBRs across the European Statistical System (ESS) including the central component of the EuroGroups Register (EGR) and to transform the current systems into a network of consistent SBRs so that they serve as the backbone of the production of European business statistics, both at country and European level.

ESBRs is one of the flagship projects implementing ESS Vision 2020 approved by the European Statistical System Committee (ESSC) in May 2014. The ESS Vision 2020 builds upon a holistic approach to reach quality and efficiency gains. It elaborates the European approach to official statistics by:

- embracing opportunities provided by the digital transformation and emerging data sources;
- putting quality as an overarching dimension of the statistical production process;
- suggesting new modes of collaboration and emphasising the importance of dissemination and user engagement to drive continuous improvements.

The ESS Vision 2020 has identified five key areas for delivering the Vision:

1. Identifying user needs and cooperation with stakeholders;

2. Quality of European Statistics;
3. New data sources;
4. Efficiency and robust statistical processes;
5. Dissemination and communication.

ESBRs contributes to the ESS Vision objectives for Quality of European statistics, New data sources and Efficiency and robust statistical processes.

The ESBRs project builds upon previous MEETS projects (mainly 'EGR' and 'Profiling', but also some aspects of 'Consistency') and integrates them in the perspective of SBRs modernisation and interoperability.

The future ESBRs system will be built on the key functions supported by EGR and the IPT (Interactive Profiling Tool) application currently under development.

1.2. The EGR and its components

The EGR is a network of registers consisting of a central register at Eurostat and local registers in the EU Member States[1] and EFTA countries[2] containing data about multinational enterprise groups (MNEs). It stores and integrates data on MNEs collected by Eurostat from National Statistical Institutes (NSIs) and Commercial Data Providers (CDPs) and releases them to national statistical authorities for national and European statistical purposes.

The EGR information system is being developed to provide and implement functions such a viewing data, updating/improving data, validating data, retrieving data, exchanging data between EGR and users, integrating data.

The EGR 2.0 system consists of four component applications:

- the EGR Identification Service (EGR IS)
- the EGR Foreign Affiliate Trade Statistics interface (EGR FATS)
- the EGR Core application (EGR CORE)
- the EGR Interactive Module (EGR IM)

The EGR applications and their data are hosted in a newly created Eurostat's secure environment for confidential microdata (SICON).

1.2.1. EGR IS

EGR IS is a comprehensive search engine supporting statisticians in 32 countries (EU and EFTA) in identifying legal units.

[1] Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom.

[2] Iceland, Liechtenstein, Norway, Switzerland.

1.2.2. EGR FATS

EGR FATS provides users (FATS statisticians) with a web interface to inspect the outputs of the consolidation process in terms of frames (consisting of legal units, enterprises and global enterprise groups).

1.2.3. EGR CORE

EGR CORE is the heart of EGR 2.0. It consolidates the input data arriving from 32 National Statistical Institutes (NSIs) and 2 Commercial Data Providers (CDPs) and generates the frame.

EGR CORE has 5 main components:

- The Source/Input area collects all microdata arriving from NSIs and CDPs, after their secure transfer and validation via EDAMIS and EDIT systems respectively. This process also ensures that appropriate identification of the legal units is performed, by making use of EGR IS.

- The Transformation area is reached only by Source area data that meet certain consistency rules. Inconsistent data can trigger new requests for data to the sources (NSIs or CDPs).

- The Consolidation area is triggered by the presence of data in the Transformation area. This process consolidates the legal units and their pairwise relationships into enterprise groups. This creates a provisional frame population which is stored in the Consolidation area.

- The Target/Output area contains a static picture (preliminary or final) of the statistical frame population which can be used in the national statistical production process. End-users (statisticians) can then retrieve data from these frame populations on EGR FATS.

- The Management module provides the necessary functionality for the administration of the EGR system.

1.2.4. EGR IM

The EGR Interactive Module (or EGR IM) provides users with an interactive web interface to validate data in the consolidation area of EGR CORE. It introduces the idea of collaborative statistical production among statisticians in 32 countries.

1.2.5. IPT

The IPT application is going to support Eurostat and the national statistical authorities in their Profiling activities concerning large MNEs.

The IPT application is an independent application. It is being developed as a prototype and it will be integrated into the universe of EGR applications.

1.2.6. State of play

EGR 2.0 is expected to be completed by summer 2016. Currently, EGR IS and EGR FATS are used in production, while most of EGR CORE’s components have been completed.

The project team is striving to put a working version of EGR CORE and EGR IM in production by the end of 2015, so that EGR 2.0 can be used in statistical production already from January 2016.

The IPT prototype will be completed by summer 2016. A working version is expected in production by March 2016.

1.3. The ESS Enterprise Architecture Reference Framework

In order to achieve its objectives of sharing, where appropriate, knowledge, methodology, tools, data, services and resources across the ESS, the ESS Vision 2020 explicitly stressed the need to develop an ESS Enterprise Architecture (EA) to enable a systematic and coherent approach to the modernisation of European statistics.

In response to this need, the ESS set up a Task Force on EA with the objective to develop a reference ESS EA framework. The Task Force concluded its work in August 2015 by releasing two documents: the ESS EA Reference Framework (ESS EA RF) and the Statistical Production Reference Architecture (SPRA). They contain a series of architectural artefacts, namely:

1) a business capability model;
2) a list of architecture building blocks;
3) a reference list of statistical services;
4) a list of enterprise architecture principles and
5) systematic references to relevant implementation standards.

They form a framework that:

- Establishes a common language to bridge business and IT across the Vision implementation portfolio;
- Provides guidance in the design and execution of ESS.VIP projects to ensure a coherent implementation of the ESS 2020 Vision;
- Allows identifying and taking advantage of opportunities for the re-use and/or sharing of available IT components;
- Supports strategic decision-making on which areas should be targeted for increased standardisation and/or interoperability;

Now we illustrate how the first three artefacts can be used to analyse the potential impact of the EGR system as part of the EBRS project:

- The business capability model describes what the ESS does and what it will need to do differently (in response to the strategic challenges and opportunities and priorities identified in the Vision). A capability is supported by adequate combination of people, skills, process, systems and technologies, methods and standards. A business capability map provides a synthetic view of an enterprise/system that can be understood by both business and IT.
- The ESS Building Blocks are homogeneous high-level functional components supporting the statistical value chain and deemed essential to realise the Vision goals.
- The reference list of business services describes about 70 generic services to be exposed by ESS Building Blocks using a SOA architectural style. The services implemented by the ESS in compliance with the Common Statistical Production Architecture (CSPA) guidelines will eventually be referenced in the Global Catalogue sponsored by the High Level Group for the Modernisation of Official

Statistics (HLG-MOS) and hosted by Eurostat. This catalogue will contain the list of IT services available for sharing or to be developed through international cooperation.

2. EA ANALYSIS OF THE EGR SYSTEMS

2.1. EGR capabilities

The EGR systems implement several business capabilities. These are supported by a chosen combination of human resources, process and governance, IT, information standards, methodology (see Table 1). They express in a business language the contribution of the EGR to ESBRs..

EGR capabilities	**1. Provide unique identification to MNEs**	**2. Integrate data on MNEs from different sources**	**3. Collaborative case management for statistical production**	**Release data and provide access to ESS Statistician**	**Profiling of MNEs**
Human resources	Central team maintaining the reference database and link with external provider data service	EGR Process manager	BR experts distributed in NSIs	Central team Master data manager	Expert in profiling distributed in NSIs
Process	Single service	Integration process and data storage	Case management process (rules and responsibilities)	Notification process of update and release	Contact HQ of group, on site visit
IT	EGR IS	EGR CORE	EGR IM	EGR FATS	IPT
Information Standards	Standard input / output Web service		Human (web) user interface	SDMX IM Web user interface	-
Methods	Fuzzy search Business rules for handling priorities	Business rules for priorities	Case management rules		Harmonised methodology for profiling
ESS capability (level 2)	Statistical Register Management	*Statistical Data Preparation and Management*	-	*Flexible data access provisioning*	Statistical Register management

Table 1. ESBRs / EGR capabilities

The capabilities of Table 1 can be mapped onto the ESS capability model of the ESS EA RF (Table 1 and Figure 1).
The EGR system contributes to the following ESS capabilites:

- Statistical Registers Management and Collection of administrative and other sources (under Data Collection);
- Statistical Data Management (under Information Resources Management) and
- Flexible data access provisioning (under Statistical Dissemination).

All these capabilities are in scope of the ESS Vision 2020.

This mapping indicates potential positive contributions of the EGR/ESBRS to the realisation of the ESS Vision 2020.

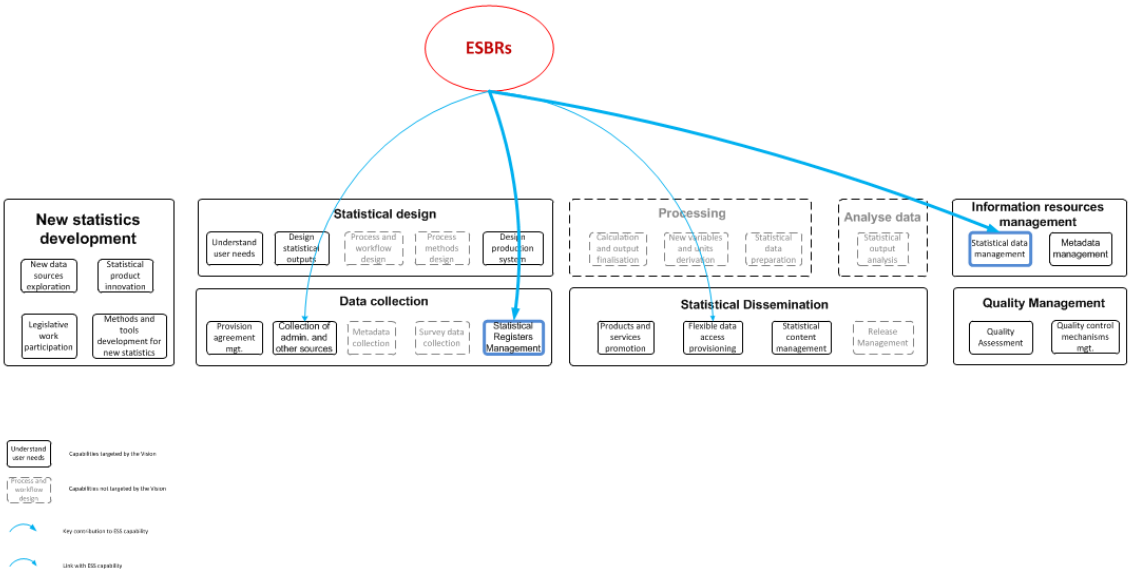

Figure 1. EGR capabilities mapped on the ESS capability model

From a portfolio/programme management perspective, this mapping allows a cartography of each project's contributions within the project portfolio enabling to identify potential synergies and overlap. Figure 2 provides a mapping of the current ESS Vision 2020 portfolio. It allows visualising potential synergies between ESBRs and project like ESDEN (European System Data Exchange Network) and SIMSTAT

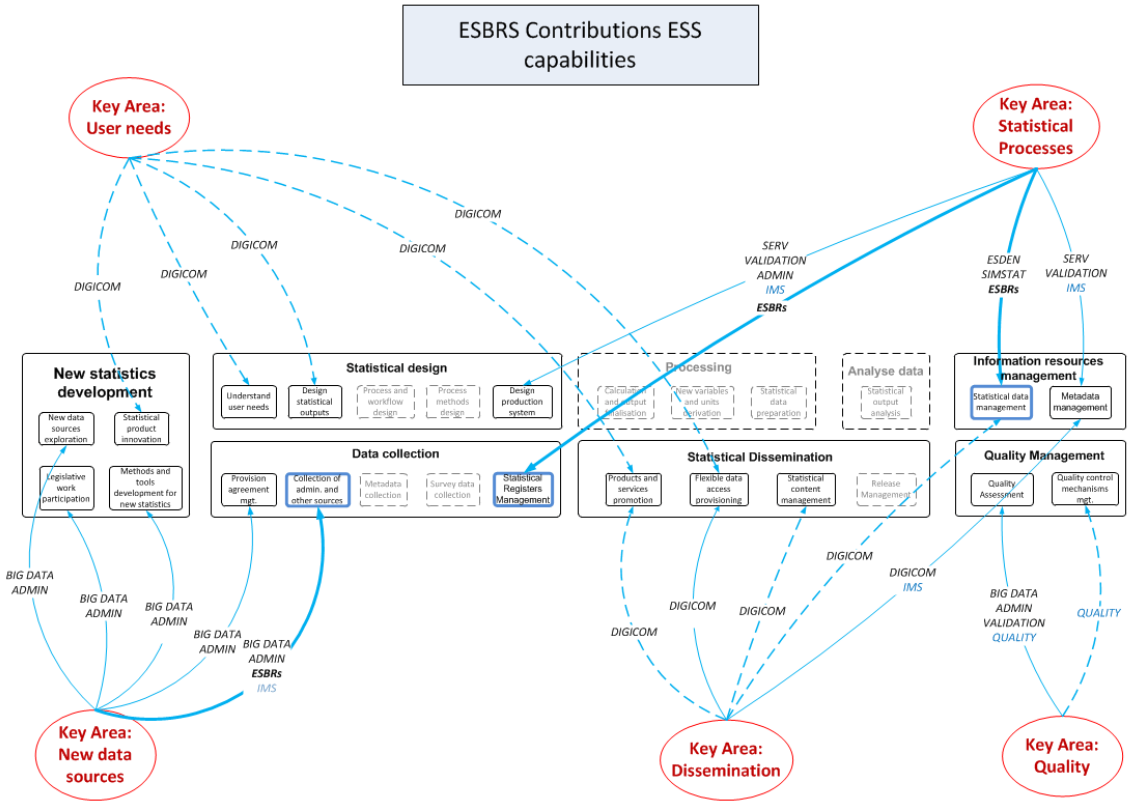

Figure 2. ESS Vision 2020 portfolio mapping and contribution of the ESBRs

2.2. EGR contributions to ESS building block and services

The list of ESS EA Building Blocks provided by the ESS EA RF allows systematically decomposing and analysing the contributions of EGR applications from a functional view and identifying which of the SPRA business service is applicable..

The Table 2 below provides the ESBRs/EGR project manager a framework to identify in collaboration with ESS Enterprise Architects and in line with the overarching goals set up in the ESS Vision 2020 which existing components could be potentially **reused** and which component should be developed **anew**.

In addition, for each building block/(*service*) which has a potential benefit for reuse across ESS members and/or across domains, a decision should be taken on the deployment approach at ESS level. The ESS EA RF identifies 4 different approaches:

- Autonomous: function/services are designed and operated without coordination with other ESS members;
- Interoperable: function/ service are coordinated through interoperability. The NSIs have the autonomy to design and operate their own functions, as long as they have the ability to exchange information and operate together effectively (through their respective information systems);
- Replicated: function/services are duplicated using de facto interoperability. The NSIs have integrated an instance of the generic function/service locally;
- Shared: function/services are common, shared and accessible to all the NSIs. There is a single instance that is shared and available to all.

In Table 2, possible deployment a scenario is illustrated using the convention colours of ESS EA RF:

Components that could be eventually developed into shared services are painted in orange colour and marked with “S”. Similarly, components that could be eventually developed into replicated services are painted in green colour and marked with “R”.

This type of decision has important consequence on the architecture of the component, with for instance shared service should be developed CSPA compliant. The additional costs should be balanced with the benefits.

EGR component / ESS building block	EGR IS	EGR CORE	EGR IM	EGR FATS	IPT
Statistical processing	**Anew** (matching) **S**	**Anew** (data integration) **R**			
Data collection	**Anew** (External data retrieval services) **S**				Anew
Collaboration			Anew		Reuse (IM)
IT Security	Reuse (SICON)	Reuse	Reuse	Reuse	Reuse

		(SICON)	(SICON)	(SICON)	(SICON)
Consumer management and tracking					
Design of collection instruments					
Identity and access management (IAM)	**Reuse** (EC authentication)	**Reuse** (EC authentication)	**Reuse** (EC authentication)	**Reuse** (EC authentication)	**Reuse** (EC authentication)
Process Orchestrator	**Reuse** (local infrastructure in NSIs)				
Dissemination				**Anew** or **reuse** (ESTAT dissemination chain components)	
Metadata Management			**Anew** (access to register metadata)	**Anew** (access to register metadata)	
Data exploration and analysis			**Anew**	**Anew**	
ESS Data exchange		**Reuse** (ESDEN)		**Reuse** (ESDEN)	
Dissemination Data Storage				**Anew** (confidential data storage for release)	**Reuse**
Primary Data Storage	**Reuse**	**Anew** (central repository for MNES) **R**			
Catalogue of reusable solutions and standards					
Quality assessment		**Reuse** (ESS validation service)	**Anew**		
Data supplier registry	**Reuse**				**Reuse**

Table 2 EGR decomposition according to ESS Building Blocks and a possible deployment scenario

The identification and deployment of statistical services is dependent on the ESBR process model which is still under development (and it is part of the Business Architecture document currently discussed by the ESBRs team). The current analysis could be then considered as mostly illustrative or inspirational.

From a portfolio management perspective, this decomposition allows consistent tracking of the developments of the target building blocks necessary to realise the ESS Vision 2020 and ensuring efficiency of investments by maximising reuse and avoiding duplication.

2.3. Global catalogue

The Common Statistical Production Architecture (CSPA) is the industry architecture for the official statistics industry, which builds on existing frameworks, notably the GSBPM and GSIM, as the necessary shared industry vocabulary. The initiative is sponsored by the HLG-MOS and managed by UNECE.

The primary aim of CSPA is to support efficient sharing and reuse of process patterns, information and services at an organization and international level.

One key requisite in achieving this goal is an ability to reliably and efficiently discover what is available for reuse to support a particular business need. This includes an ability to efficiently assess whether a potentially reusable artefact is, in fact, "fit for purpose" in practice when it comes to supporting that particular business need.

Catalogues of reusable resources have a key role within CSPA. They provide lists and descriptions of standardized artefacts, and, where relevant, information on how to obtain and use them. The catalogues can be at many levels, from global to local. For example, it is envisaged that each statistical organization will have catalogues of processes, information objects and Statistical Services.

The CSPA Global Artefact Catalogue provides information about resources and potential collaboration partners, helping to ensure that Statistical Services conform to the requirements of CSPA. The Global catalogue serves primarily the goal of the reinforced international collaboration. It the ESS context, it matches the objectives of sharing infrastructure as promoted by the ESS Vision 2020.

Governance and support mechanisms and processes are being defined to ensure the continued relevance and utility of the Global Artefact Catalogue both at Industry level and at ESS level (ESS.VIP.SERV). The Global Artefact Catalogue is hosted by Eurostat.

The Global Artefact Catalogue is relevant for many services identified in ESBRs. As an example, if a positive business case is identified at industry level, the EGR IS could be wrapped as a CPSA service to be published in the Global Artefact Catalogue.

3. CONCLUSIONS

The paper aimed to achieve two objectives:

- first, to analyse the potential contributions of the EGR systems for ESBRs and ESS modernisation;
- second, to demonstrate the relevance of architectural artefacts (as provided by the ESS EA RF) for project managers that have to design and manage complex and ambitious projects at ESS level and for portfolio managers that have to maintain coherence and ensure efficiency of the developments across projects .

The analysis indicates that EGR 2.0 to be released in production by the end of 2015 and finalised by summer 2016 will deliver essential capabilities for the ESBRs and for the ESS in general.

However, despite the fact that EGR 2.0 is highly modular and an SOA style has been widely followed in its design, EGR 2.0 cannot yet deliver fully reusable and interoperable services at ESS level in line with the ESS EA RF.. This can be explained because the EGR 2.0 was fully specified and (to a certain degree) developed before the ESS Vision 2020 was set up and the ESS EA RF released.

This advocates in favour of involving ESS Enterprise Architects as early as possible in the design of the key ESS infrastructural components. In addition, in the ESBRs context, it suggests that an architectural approach as currently followed by ESBRs combined with a solid cost benefit analysis is a valid approach for deciding of the future improvements of the EGR systems.

In conclusion, on one side, the present paper demonstrates, if needed, that the EGR has a strong potential for delivering key capabilities in the context of a more interoperable and integrated system of business registers (in line the ESS Vision and the ESS EA RF). On the other side, the ESS EA RF proves to provide a relevant framework for analysing ESBRs / EGR capabilities advocating for its systematic use for other ESS Vision 2020 implementation projects.