

**26th Meeting of the Wiesbaden Group on Business Registers
- Neuchâtel, 24 – 27 September 2018**

Enrica Morganti
Eurostat

Session 3 Integrated statistical registers systems

The Interoperability Framework of the European System of Business Registers

Introduction

One of the European Commission's ten priorities is to remove barriers to a digital single market in Europe. All EU Member States are digitising their public administrations to save time, reduce costs, increase transparency, and improve both data quality and the delivery of their output. Interoperability is the key factor in making this possible. Interoperability allows different organisations to electronically exchanging meaningful information in ways understood by all parties. On 23 March 2017 the European Commission adopted the new European Interoperability Framework (EIF) as part of the Communication (COM(2017)134) to give guidance to European public administrations on how to set up interoperability for public services.

In statistics, this policy is part of the European Statistical System (ESS) Vision 2020. Since 2013, Eurostat and the ESS partners are working on a portfolio of projects for its implementation. The European System of Interoperable Statistical Business Registers (ESBRs) is one of those projects and it is about establishing the conditions of interoperability between the national statistical business registers (SBRs) and the EuroGroups Register (EGR). Since 2008, they started exchanging cross border statistical confidential information about large and complex multinational enterprise groups. The ESBRs project objective is to increase standardisation and automatization and make data exchange more efficient. The output of the EGR is a statistical global frame containing information about multinational enterprise groups and their statistical units used by all ESS partners. If national statistical business registers (SBRs) and the EuroGroups Register (EGR) are not interoperable, the statistical frame contains inconsistent information about multinational groups and this affects the quality of statistics based on it. Ultimately, the ESBRs project aims to increasing the quality of the EGR frame and make it a reliable 'backbone' to coordinate ESS statisticians and their output.

The paper presents the ESBRs Interoperability Framework, the document describing the legal, organisational, semantic and technical aspects of interoperability as developed within the project¹.

¹ Working version available at July 2018.

Keywords: European Interoperability Framework, Statistical Business Registers, EuroGroups register (EGR), ESBRs project, EU Vision 2020, data exchanges, automation and standardization, sharing and re-using solution, European Statistical System.

About the European System of interoperable Business Registers (ESBRs)

The European System of interoperable Business Registers (ESBRs) is composed by the national statistical business registers, located and managed in the EU Member States and the EuroGroups Register (EGR), located and managed at Eurostat. They exchange statistical information about legal units and enterprises that are linked by cross border control relationships and form multinational enterprise groups. The main focus is on the multinational groups that have a significant presence in the EU. The need to exchange statistical information on multinational groups comes from the fact that each national statistical office alone is unable to derive a complete and correct picture from its national administrative sources. They can observe only a 'truncated' view of the multinational groups for the legal units that are resident on their territory and some cross border relationships, while information about not resident legal units and the control chain outside its territory is usually not reachable.

Since 2008 the national statistical offices of European Member States and EFTA countries have started to cooperate pooling together micro data on legal units, control relationships, enterprises and enterprise groups exchanging data in the EuroGroups Register (EGR). The EGR receives information from the countries, consolidates it into a global frame and disseminates it back. The EGR global frame provides the global view of the multinational groups and their enterprises, legal units (identification, demographic and economic characteristics), relationships and control, shareholdings of at least 10%. Among the group's characteristics, the Ultimate Controlling Institutional Unit (UCI) is the key one for FATS. The UCI is the institutional unit, proceeding up a foreign affiliate's chain of control, which is not controlled by another institutional unit². To ensure the quality of FATS data, statisticians in different countries should apply the same methodology and coordinate among themselves on the UCIs to consistently define their Inward or Outward populations, launch data collections and publish FATS data. The EGR database and the global frames are restricted to users of national statistical offices, national central banks and the ECB.

In 2013 Eurostat and the ESS partners agreed to work on the European System of Interoperable Statistical Business Registers (ESBRs) projects to increase the interoperability between the national statistical business registers (SBRs) and the EuroGroups Register (EGR) and improve the process and organization for creating the global frames and its quality.

EU Members States have actively contributed to the project deliverables: an ESSnet has worked from 2013 to 2016 delivering important elements of the ESBRs at technical level, including designing and testing of EGR 2.0 modules, that were finalised in 2016 and allowed significant improvements in the quality of the global frames. The ESSnet also contributed preparing methodological parts of the Global profiling and drafting the ESBRs business architecture and interoperability framework. Since 2016, dedicated Task Forces on Global profiling methodology and on ESBRs business architecture and interoperability framework are working to finalise the results achieved by the ESSnet. EU Member States are also very active in testing the

² REGULATION (EC) No 716/2007 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 June 2007 on Community statistics on the structure and activity of foreign affiliates.

methodologies and the tools developed by the ESBRS project. The Commission is financing every year individual or multi-beneficiary grants to test profiling and collaborative activities as well as pilots on interoperability). FATS users have also been involved in the identification of user requirements and in the alignment of calendars and timing of the global frames. Their active participation is a crucial to the success of the project.

The project will end in 2020 and by then the target objective is to deliver global frames with a fit-for-purpose quality level that is accepted by FATS statisticians to use it as 'backbone' for their data production.

The European Interoperability standards

There are two document setting up the standards for interoperability in the EU: 1) the New European Interoperability Framework (EIF), part of the European Commission Communication (COM(2017)134) of 23 March 2017, setting up standards for the logical descriptions of interoperability models and 2) the European Interoperability Reference Architecture (EIRA)© setting up standards for architectural solutions. The ESBRS business architecture and interoperability framework follow to such standards.

Target audience of the ESBRS IF

The ESBRS IF serves two target audiences in the project: statisticians and architects. It is a challenge to have one single document serving such different target audiences. As a compromise solution, the core document of the ESBRS IF describes the interoperability according to the four layers proposed by the EIF and it is directed to the audience of statisticians, while the Appendix contains the architectural views according to EIRA© standards and is directed to the audience of architects.

The target statisticians can be producers of the ESBRS outcome or users. They read the ESBRS IF to understand the process, the methodology, the concepts and data used, the organizational model and the data exchange modalities. The business process is described for them using the Business Process Model Notation (BPMN). They can recognize their role and the interactions with the other partners in the ESBRS.

The target architects can use the standard EIRA© views to understand the ESBRS Architecture Building Blocks (ABBs), benchmark their own existing solution, do gap analysis and design architectural solutions (Solution Building Blocks (SBBs)) to improve their level of interoperability in the ESBRS.

The interoperability layers according to the New European Interoperability Framework (EIF)

1. Legal layer

In the European Union each public administration contributing to the ESBRS works within its own national legal framework. Legal interoperability in the ESBRS is about ensuring that organisations operating under different legal frameworks, policies and strategies are able to work together.

The EU Council Regulation 177/2008 of 20 February 2008 for establishing a common framework for business registers for statistical purposes is the main legal text for the data exchanges that are used in the ESBRs and a basis for the legal interoperability. The transmission of confidential data is limited to the extent that it is necessary for the production of specific statistics in the European Union and has to respect specific rules and limitations.

Council Regulation 177/2008 (Article 10) defines the limits of data exchange. In particular, the exchange of confidential data has to be done exclusively for statistical purposes and in accordance with national legislations. Article 11 *allows* the exchange of confidential data between the appropriate national authorities of different Member States, but it is not a mandatory provision. Exchange of confidential data is functional to ensure the (good) quality of the multinational enterprise group information in the European Union. National central banks may be party to the exchange in accordance with national legislation.

Council Regulation 177/2008 (Article 11) requires something more. Appropriate national authorities of different Member States *shall transmit* data on multinational enterprise groups and their constituent units to Eurostat and Eurostat, to ensure a consistent record of data, transmits back to each Member State data on a multinational enterprise group. A territorial limitation applies her. Consolidated data on multinational groups, including records coming from other countries, can be transmitted to Member State only *when at least one legal unit* of the group is located in the territory of that Member State.

Finally, Council Regulation 177/2008 Article 12 *allows* the exchange of confidential data, exclusively for statistical purposes, between Eurostat and national central banks, and between Eurostat and the European Central Bank. The exchange has to be *explicitly authorised* by the appropriate national authority.

As seen, the exchange of confidential data constitutes a string basis for the legal interoperability, but it is subject to limitations to ensure that the statistical confidentiality is guaranteed and that data exchanged are carried out for *statistical purposes only*. The format, security, confidentiality measures, and procedure for the transmission of data are defined respectively in the Commission Regulation (EC) No 192/2009 (implementing article 11) and in the Commission Regulation (EC) No 1097/2010 (implementing article 12).

All parties in the data exchanges (the Commission (Eurostat), national authorities, national central banks and the European Central Bank) receiving confidential data pursuant to Articles 10, 11 and 12 shall treat information confidentially in accordance with Regulation (EC) No 223/2009.

The Framework Regulation Integrating Business Statistics (FRIBS) will in future repeal Regulation 177/2008 and the Commission Regulation.

In addition to the binding legal instruments on data exchange, the European Statistical System (ESS) Vision 2020, endorsed by the European Statistical System Committee (ESSC) in 2014 and in particular the ESBRs project, are based on agreed principles of collaboration. ESS members have recognized that exchange of micro data is an essential instrument to bring advantages to everyone in those statistical domains related to globalisation.

2. Organizational layer

The chain management model

At the core of the ESBRS organisational interoperability there is the ESBRS chain management model. Chain management is a way of working together in a non-hierarchical way (ref. Gert Buiten). The concept suits for the ESBRS where independent organisations (NSIs, and Eurostat – and in future possibly the NCBs) aim to work together to achieve common goals without being a hierarchical organisation. The increasing organizational complexity of large and complex multinational enterprise groups has an impact on the quality of business statistics produced by different countries in the ESS. Different producers or compilers manage individual processes that have dependencies one to another and that are linked in a whole chain. None of the individual alone can oversee the system of dependencies and control the quality of the output. Chain management is necessary to connect the various processes as part of a chain and to ensure that decisions or changes executed by each partner are taken having in mind the benefit of the whole chain. This avoid that inconsistencies in the output are detected at a later stage and reduces the costs to correct them.

Chain management considers the complete chain and the interdependencies between the different parts of the chain but also relies on the quality of the single processes themselves. It is the result of different synergies. To achieve a good chain management it is necessary first to optimize the individual processes themselves and then to optimize the chain as a whole. The goals of chain management are to improve the quality of the output and of the chain as a whole.

In order to have a successful chain management the partners have to agree on a certain set of principles, roles and responsibilities. The chain manager has the role dealing with the tactical and operational level and is responsible for the output of the chain. Every partner in the chain has the responsibility to coordinate with the others for the good performance of the whole chain. This could regard the methodology, the calendar and timing, the data validation etc.

The ESBRS chain is complicated because of the high number of (national) partners and their dependencies. For 2020, the objective is to set up a ESBRS chain management corresponding to the smallest possible set of users and focusing on the output for FATS statistics. The ESBRS smallest chain is: NBR→Global profiling→EGR→ FATS.

1. NBR National Business Registers manage the process 'Produce national frame'. This includes the maintenance of the 'live' national statistical business register and the activity of 'National profiling'. The global part of national frame is input for 'Produce global frame'.

2. EGR The process 'Produce Global Frame' is managed by Eurostat. The output is the GF (initial and final) containing the information on GEGs active within EU (+EFTA) area at a certain reference year (T), their constituent LEUs, relationships REL and enterprises ENTs.

3. Global Profiling The process Global Profiling process delineates the enterprises across border using the concept of 'global enterprises'. This serves to guide countries achieving consistency when delineating the national statistical units used for the data collection.

4. FATS statistics measure the structure and activity of foreign affiliates (FATS). FATS data can be used to monitor the effectiveness of the internal market and the integration of EU economies in the global market. **Inward FATS** are statistics measuring the structure and activity of enterprises in EU countries that are controlled by multinational groups with UCI outside the EU.

Outward FATS are statistics measuring the structure and activity enterprises outside EU that controlled by multinational groups with UCI in EU countries.

Two versions of the global frames

To realize the chain management, the ESBRS IF creates (versions of) standardized global frames and delivers them to users according to their needs. By 2020 ESBRS global frames are to serve FATS statistics as 'backbone'. Thus it is necessary to have a frame methodology to coordinate the quality of the national frames that are input for the ESBRS global frames. The ESBRS is designed to produce (at least) two global frames initial and final to serve FATS statistics.

The initial global frame has the role of a timely 'survey frame' for FATS in the phase 'Create frame & select sample'. The initial global frame serves FATS processes based on surveys. It does not serve FATS processes based on administrative data only.

Several countries produce Outward FATS statistics using direct surveys and other using administrative data (in connection with FDI statistics). Only the first group needs the initial global frame in the phase 'Create frame & select sample'³. Both groups will need the final frame in the phase 'Calculate weights'⁴. Regarding Inward FATS, most of the countries produce them as a subset of SBS. All economic data are collected from the SBS survey populations and the UCIs for the FATS breakdowns are collected separately from various sources. Those using direct surveys also need the initial global frame (like for Outward FATS), otherwise they need only the final global frame in the phase 'Calculate weights'.

For producing the initial global frame, it is used the final frame (T-1) + the Entries in (T) - the Exits in (T), and the updating of the characteristics at (T). In practice the Entries in (T) may be very limited, because to determine the Entries in (T) in the global frame it is not only necessary to know the births, but also the relationships to a group and usually they are available only at a later stage. The Exits (T) i.e. ceased and liquidated units in (T) can be known more easily. Also, some characteristics relevant to update the frame can be available, first of all the UCIs that have changed during year (T).

The process 'Producing global frame' initial for the reference year (T) is similar but not the same as the process 'Producing global frame' final for the reference year (T). The differences are followings:

1. On a voluntary basis NSIs send an initial GP NF (T) by March (T+1);
2. GP NF contains only updated records, even if the data structure is the same;
3. GP NF contains updated records for LEU, GEG or REL (T) depending on data availability in the countries.
4. Information on changed UCIs of global groups for the reference year (T) could become available during the process 'Global profiling' on the reference year (T-1) and be input for 'Produce global frame' initial one.

³ (GSBPM phase 4.1)

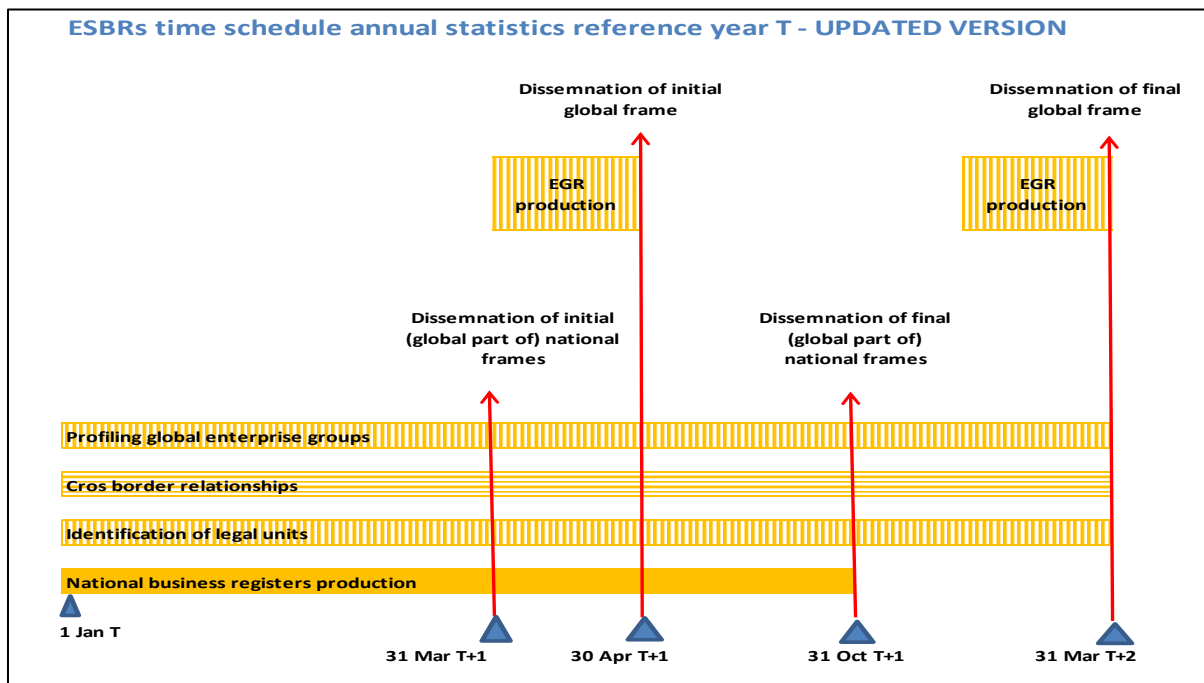
⁴ (GSBPM phase 5.6)

5. Eurostat integrate the initial GP NF (T) and the results of 'Global profiling (T)' with the final global frame of the previous year (T-1) and disseminates the initial global (T) frame by April T+1⁵.
6. The initial global frame is input for FATS statisticians in the phase 'Create frame & select sample'.
7. The initial global frame is input for the process 'Global profiling' on the reference year (T)

A final global frame for reference year T is disseminated in March (T+2). The final global frame has the backbone role to coordinate FATS statistics and their UCIs to ensure EU consistency. The final global frame serves FATS processes during the validation and analysis of statistical output and it is needed in the phase of 'Calculate weights'⁶ (GSBPM phase 5.6). All FATS processes in the ESS coordinate on it to ensure the best quality and consistency.

The production of the final frame for reference year (T) is the following:

1. NSIs send an final GP NF (T) by October (T+1);
2. GP NF contains complete records for LEU, GEG, REL and ENT (T);
3. Eurostat integrate the final GP NF (T) and the output of the process 'Global profiling' (T) and disseminates a preliminary final global (T) frame for validation activities;
4. NSIs can continue to update the ENT (T) until March (T+2);
5. Eurostat disseminates a final global frame (T) by March (T+2).
6. The final global frame is input for FATS statisticians in the phase 'Calculate weights'.
7. The output of process 'Global profiling' on the reference year (T) and be input for 'Produce global frame' (T) final one.



⁵ For some countries April year (T+1) is late, because they need to pre-inform respondent about an incoming survey by the end of the reference period. For the moment this is the best possible compromise, based on different situation in the countries.

The ESBRS processes describe the data exchanges between national business registers and EGR, profiling and it is based on FATS user needs. The description of the processes and sub-processes use the Business Process Model Notation (BPMN) to read the models and connect them to the respective concepts and data structures.

The ESBRS process consists of two main processes:

- 1) 'Produce national frame'.
- 2) 'Produce global frame'.

The output from 'Produce national frame' is input to 'Produce Global Frame'. The output from 'Produce Global Frame' is the initial and final global frames on multinational enterprise groups, their constituent legal units and enterprises. The core content of global frame is: Global enterprise groups (GF GEG), Enterprises (GF ENT), Legal units (GF LEU), Control relationships between legal units (GF REL), Relationships between enterprises and legal units (GF LEL).

The two main processes are supported by three sub-processes:

- 3.1) 'Identify legal unit'.
- 3.2) 'Validation of cross-border relationship'.
- 3.3) 'Profiling of global enterprise group'.

The process 'Identify legal units' has a crucial role for interoperability in the ESBRS. It enables the integration of information from different sources into the EGR⁷. Relationships between parents and subsidiaries legal units serve to create the multinational groups structures. The sources of information differ between countries and some have access to information earlier than others, thus sharing the, with a partner countries allows improving timeliness and efficiency of the process. For the largest and most relevant multinational groups the correctness of the multinational groups' information is essential to ensure good quality of FATS statistics and 'Global profiling' is the collaborative process allowing different national teams to cooperate and share statistical information useful for the validation of the multinational groups and their statistical units.

The ESBRS process is very complex and organized in several data exchanges and activities that run all over the year and involve the participation of 32 EU Member States and EFTA countries, Eurostat and experts from different statistical domains in the ESBRS chain management (business registers, profiling, FATS). It is not possible to describe here in detail all the steps of the ESBRS process, but the main one.

The main complexity of the ESBRS process derives from the fact that the input comes from different national processes and interoperability solutions have to take into consideration their timing, calendars, sources and methods. The ESBRS is not about changing the national process, but about harmonizing and standardizing their output (by aligning calendars and methods) to ensure that each part contributes in a consistent manner to deliver a qualitative output for users.

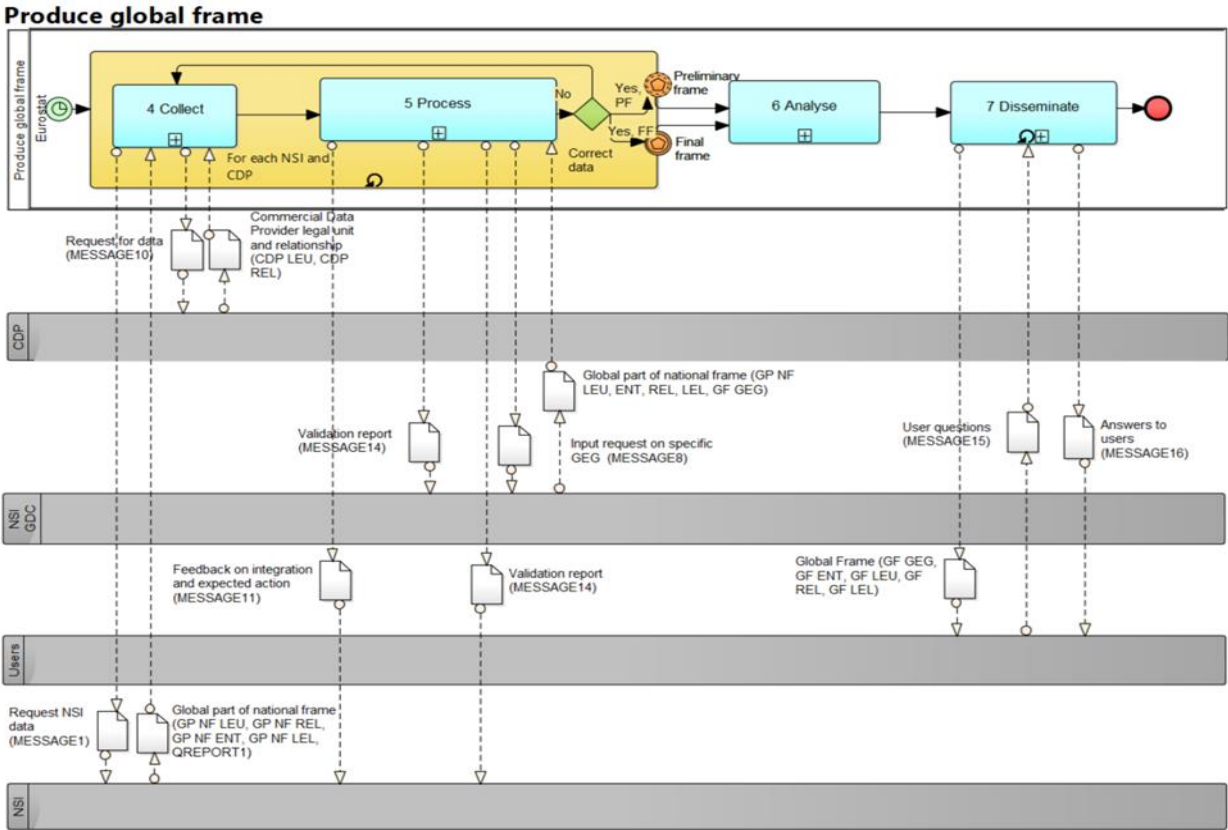
Next, the overall ESBRS process is shown in the picture 'Produce global frame'. The description of the different sub-processes are part of the ESBRS interoperability framework..

⁷ It can also facilitate the linkage between TEC, STEC and FDI reporting units and ESBRS legal units. Also it offers the maintenance of links with other relevant administrative identifiers.

Data exchange mechanisms

Another relevant aspect of the organizational layer of interoperability is the data exchanges mechanism used in the ESBRs. There are several mechanisms that can be utilized between the national statistical offices and Eurostat. The ESBRs project is financing a series of interoperability Pilots to allow EU Member States to test the standardized and automated data exchanges. The pilots will run until middle 2019.

According to the interoperability Pilots, data exchange channels agreed by the parties are based on SDMX format. The conversion of CSV files into the SDMX-ML format has to be carried out using the SDMX Converter. The SDMX Converter tool, the necessary SDMX definitions, as well as support for the configuration are provided by Eurostat. NSIs in the ESBRs can also decide to set up an automatic generation of their output files in SDMX-ML format and submit them as input to EGR by using existing tools (e.g. SDMX-RI). The extraction of output files from national statistical business registers for submission to EGR can be established via a dynamic data exchange where the EGR pulls data from a data set exposed by the responsible NSI. This pulling mechanism provides the maximum level of burden reduction, standardization and automation, as well as increased error avoidance. NSI can reuse existing tools (e.g. SDMX-RI pull-mode), infrastructure (e.g. ESDEN) and applications (e.g. EGR, national systems). Finally national statistical business registers and EGR could dynamically communicate and exchange data without human intervention by means of an interface for machine-to-machine interaction in the form of a "web service".



3. Semantic layer

The Semantic layer is the most important layer supporting interoperability and data exchanges between different organisations. It enables actors in the ESBRS chain to communicate in a meaningful manner. The following components form the semantic layer: Concepts and Data Structures.

The Concept describe what is to be measured in the ESBRS output (the global frames). The variables indicate how the concepts are measured and the represented variable how they are represented in the output. Concepts, variables and represented variables refer to certain Populations and Unit types that are relevant for users and on which measurements are made.

The component of the ESBRS data structures are the data sets, the data structures and the variables. Data set is the input or output of a process step, or a data transmission between actors in the ESBRS process. It describes a population of statistical units of one or more types and relationships. Each data set is named to ease referencing them in the documentation. Each data set is composed of one or more data structures. The data structures consist of variables on a specific type of statistical unit or relationship. The variable naming follows the conventions laid out in the Concepts chapter. In more practical terms, one data set consists of one or more tables of data. Each table is dedicated to a specific statistical unit and therefore will have its own data structure.

A crucial requirement for successful interoperability is the unique identification of data sets. Each data structure is identified using Data set identifier, Frame reference year, Frame population version, Date of validity referring to the transmission. To achieve and maintain this unique identification of data sets, all four variables are required in every data exchange between NSI and EGR.

4. Technical layer – Applications

The ESBRS application system is composed of four EGR 2.0 integrated applications and one additional not integrated application used for the activity of Global profiling. By the end of 2020 also this application will be integrated.

The four integrated applications are:

- The EGR Identification Service, a module to identify legal units and assigns unique identification numbers to these units. The legal entity identifier number (LEID number) is the unique identification number assigned by the EGR Identification Service. Access to the EGR Identification Service is restricted to users at national statistical institutes of the EU and EFTA countries.
- The EGR Core, a module for the EGR administrators at Eurostat. EGR CORE is the interface where the EGR administrators can monitor and execute the EGR processes (Monitoring the data files and data in the EGR input area and transformation area, Starting the consolidation process, Generating EGR frames, Monitoring the data of the EGR consolidation area, Managing user accounts of the EGR modules)
- The EGR Interactive Module, a module to allow business registers staff in EU Member States to browse and modify data in the consolidation area of the EGR database. With EGR Interactive

Module it is possible to modify the group structures, the relationships between group members and the characteristics of the enterprise groups' data during the EGR production process.

- The EGR FATS interface, a module allowing FATS users to browse, compare and download data on the initial and final populations of the EGR. Users can browse the EGR data, download data files in predefined formats and compare the different annual EGR populations with statistical tables.

- The Interactive Profiling Tool (IPT) is the on line application used NSIs in their Profiling activities concerning large MNEs. It is not currently integrated with the EGR modules. An on-going project is analyzing and implementing a short term and a long term solution for integration.

5. Governance and challenges to realize the ESBRs interoperability framework

In 2013 Eurostat and the ESS partners agreed to work on the European System of Interoperable Statistical Business Registers (ESBRs) projects to increase the interoperability between the national statistical business registers (SBRs) and the EuroGroups Register (EGR) and to improve the process and organization for creating the global frames and its quality.

The ownership and management of the project is assigned to Eurostat that reports to the ESS partners' governance bodies: at strategic level to the Business Statistics Director Groups that approves and endorses the projects deliverables, at operational level to the Working Group Business Registers and Statistical Units that discussed the deliverables and agree on technical. The ESBRs Steering Groups, composed of a representation of EU Members States, steers the project and provide strategic guidance and orientation to the project manager.

By 2020 the achieved level of interoperability will have to be benchmarked and evaluated by the ESS partners. The results of the interoperability Pilots will provide evidence about the different situations in the ESS partners, the feasibility of the interoperability solutions and the way forwards to implementation.

Reaching the ESBRs chain management will be one of the main challenges ahead in the final years of the project. The complete chain and the interdependencies between the different parts of the chain have to be agreed. Achieving the quality of the single processes themselves is a pre-condition. Then, national statistical business registers (SBRs) and the EuroGroups Register (EGR) can become more interoperable. The ESBRs interoperability can contribute achieving the 'backbone' of the global frames and better coordinate ESS statisticians and their statistical output.