Joint UNECE/OECD/Eurostat Meeting of the Group of Experts on Business Registers
14-15 September 2011, OECD, Paris, France   *Dominique FRANCOZ*   *Zsolt VÖLFINGER*   *Eurostat*   **Item 3**   *Linking statistical business registers across agencies, statistical domains and among countries*
The EGR version 2: an improved way of sharing information   ***on multinational enterprise groups***

Introduction

The EuroGroups Register (EGR) is foreseen to become the platform that supports the production of micro-based statistics on globalisation in Europe, both on country and European level.

The EGR system was settled down in 2009 (reference year 2008) and is now performing its third cycle on reference year 2010 with the contribution of business register staff in the National Statistical Institutes (NSIs) of the EU and EFTA countries.

One of the main objectives of EGR version 2 (V2) is enabling statisticians and register staff of NSIs and NCBs to consult, to retrieve and to update EGR data. In the meanwhile, lessons were drawn from the experience of the first cycles in order to improve significantly the supply of the database.

The EGR will be a common tool within the European Statistical System and, accordingly, in line with the vision for statistics adopted by the Commission.

This paper is aimed to present the principles, new features and tools of the EGR V2 with a special attention on the quality issues and the international identification of legal units. Indeed, the EGR, if extensively used, could help to enhance the quality of statistics on globalisation by acting on coherence between statistical areas (FATS, FDI) and between countries, and by proposing common concepts and definitions. However, the EGR itself could be analysed under the light of quality in statistics. On that regard, the users have a role to play in the improvement of the quality of the register.

PART I - The transcription of the principles of quality in the context of the EGR

Quality in statistics is an important concern of the European Statistical System. The quality criteria are included as principles in the European statistics Code of Practice that is the result of a common agreement between the National Statistical Institutes and Eurostat. They are detailed in the ESS handbook for quality report (Eurostat publication – 2009)[1]. These criteria generally apply to the output of the statistical process. However, it is possible to give an interpretation of them in the context of the EGR.

1- Relevance:
Relevance is defined, in the ESS handbook, as "*the degree to which statistical outputs meet current and potential user needs. It depends on whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.,) reflect user needs.*"

This criteria fits with the EGR context since it is foreseen to fulfil needs of users involved in statistics on globalisation. The main users are the FATS and FDI statisticians, but also the statisticians working on profiling multinational and complex groups[2].
What should be ensured, in that regard, is the practical utility of the EGR: is it is useful for users? Does it cover their needs?

Moreover, the concepts developed in the EGR should be agreed with the users, for example the definition of the ultimate controlling institutional unit (UCI) or the special purpose entities (SPEs). The EGR would then be a powerful tool if it would provide to the users common definitions, concepts and information on multinational enterprise groups.

Several steps in that direction have already been achieved since a new definition of the UCI has been proposed commonly between FATS and EGR statisticians. On the same way, a task force (which brought together FDI, FATS and EGR statisticians) which was targeted to define FDI needs resulted in a proposal to identify special purpose entities in the EGR.

2- Accuracy and reliability
*"The **accuracy** of statistical outputs in the general statistical sense is the degree of closeness of estimates to the true values."*

In the case of the EGR, this definition should be adapted. It has also to deal with reliability. It means that the information on the MNEs included in the EGR should be close to the true situation. Several elements need to be considered to fulfil this criterion: the EGR should include the frame populations of the statistical users (and maybe first the outward FATS needs); there should not be under-coverage (for example, missing legal units in a given MNE, missing MNEs according to the target population) or over-coverage (legal units registered several times under different identification numbers); the elements necessary to delineate the frame populations should be reliable in order to not misclassify a MNE in one population or another (for example, the location of the UCI is crucial to attribute the MNE to the inward FATS or the outward FATS population).

---

[1] The definitions given in the document are extracted from the ESS Handbook for quality report
[2] Currently, the actions on profiling are undertaken in the frame of an ESSnet that groups 7 NSIs. In the future, profiling of multinational enterprise should be enlarged to all the EU and EFTA NSIs.

3- Timeliness and punctuality

The "***timeliness*** *of statistical outputs is the length of time between the event or phenomenon they describe and their availability.*
Punctuality *is the time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by Regulations or previously agreed among partners."*

This element is also a crucial one for the EGR: the data base is intended to be used as a frame population for surveys (FATS and FDI). It means that the information included in the EGR should be up-to-date as much as possible. The requirements of the different users in that regard are not necessarily the same and should be fulfilled on the best possible way. The EGR could be used for sample purposes but also for imputation purposes at the end of the process. The updating schedules of information on EGR are not the same in the different countries. So, in order to propose an up-to-date register, it is intended to not wait until the last country has sent its data to disseminate the EGR, but to propose several versions that will gradually integrate the updating of the different participating NSIs. On that aspect, the users have a role to play that will be detailed below.

4- Accessibility and clarity

"Accessibility and clarity refer to the simplicity and ease with which users can access statistics, with the appropriate supporting information and assistance."

The accessibility principle has several declinations in the EGR: on a regulatory point of view, exchanges of micro-data information have been defined in several regulations; on a pragmatic point of view the content of the EGR should be released to the users via a web platform.

The clarity should be ensured by using concepts and definitions that are common to the producers and the statistical users. Moreover, recommendations manual and guidelines can support the use of the database.

5- Coherence and comparability

"*The **coherence** of two or more statistical outputs refers to the degree to which the statistical processes by which they were generated used the same concepts - classifications, definitions, and target populations – and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly. Examples of joint use are where the statistical outputs refer to the same population, reference period and region but comprise different sets of data items (say, employment data and production data) or where they comprise the same data items (say, employment data) but for different reference periods, regions, or other domains. **Comparability** is a special case of coherence and refers to the latter example above where the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains."*

The transcription of the coherence criteria into the EGR is not straightforward, since the EGR is intended to improve the coherence between statistics on globalisation by proposing unique information to the users. The information in the EGR is then supposed to be unique and coherent. However, the coherence is a concern in the building of the EGR while several countries may bring information on the same MNE and on the same financial relationship in case of cross-border relationships. Priority rules have been settled down to treat potential

conflicting cases. Moreover the EGR V2, with the authentic source principle, intends to reduce the possibilities of incoherencies.

The comparability aspect is approached in the EGR through the concern of having information comparable over time.

PART II - The EGR version 2 and its principles

The first version of EGR is now performing its third yearly cycle. The present EGR process is based on information from commercial data providers. This information is submitted to the NSIs of the EU and EFTA countries to verify the delivered legal units, relationships and group structures. Several exchanges of information between the NSIs and Eurostat take place along a yearly-process. This process is very work-intensive for the NSIs, especially for them which already have a database on MNEs.

The EGR V2 is intended to improve the quality of the EGR, not only regarding the output, but also regarding the process. It should reduce the burden for the participating NSIs. This chapter will present the main features and their impact on the global quality of the EGR.

1- Authentic source and authentic store concept
EGR V2 will follow the concept of authenticity in storing and identifying data. In EGR V2 the data will be stored in an authentic store and one specific source of the data will be marked as authentic source.

Authentic store means the place where the original and latest version of a data element or a set of data elements is stored. Authentic source means the source of which data are used as the 'truth' without additional investigation. A source is appointed as authentic if it meets certain quality requirements, e.g. comprehensiveness, accuracy, reliability, timeliness, punctuality and comparability. Data defined/validated by an authentic source can not be overwritten by another source. Even if values are missing no other sources will be used to provide these.

In the EGR network the authentic store for national units is managed by the NSI, which implicitly means that the NSI is the authentic source for national units of EGR. A NSI may decide within their national network to define subsystems of authentic stores and sources, e.g. with National Banks it could be agreed that the authentic store for financial institutions is the RIAD register.

In the EGR V2, for national legal units and national enterprise units the authentic store will be the NSI, for non-EU units the authentic store will be the EGR. For cross-border relationships the authentic store will be EGR, the authentic source can be NSI of the parent company or NSI of the subsidiary or NSI of the UCI. Availability of relationship data differs from country to country; the final solution applied in EGR should be defined in the future.

The aim of the concept is that for all authentic entities provided by NSIs the final EGR population frame of a reference period should be an exact copy of the picture in the national statistical business registers of the NSIs. The EGR considers these data as validated data by definition.

Regarding national data for which the NSI is not the authentic source the EGR will process and consolidate data from different sources which (as in the present situation) will be sent to

the NSIs for updating and validation. With this process the NSI can receive data (e.g. cross-border relationships) from commercial data providers for the creation of their national enterprise group register. Regarding data for which the EGR is authentic store and a NSI or NCB makes the choice to store these data also in the national environment the recommended situation is that the NSIs and NCBs keep an exact copy and put a maintenance procedure in place to update this copy regularly.

With the concept of authenticity the roles of central EGR and national registers will be precisely defined; the enterprise group information stored in national business registers and in EGR will be identical at the time of the statistical outputs. The new concept could retrieve the authentic national information from national registers to EGR and could reduce the burden for the participating NSIs.

2- Identification service

In the implementation of the concept of authenticity the identification service for legal units is a critical factor. This service enables statisticians and register staff in NSIs and NCBs to identify foreign legal units in the EU as well as outside the EU.

In the present EGR the identification of foreign units is one of the major challenges. NSIs usually do not have complete information on foreign legal units, e.g. on foreign parents, foreign subsidiaries; they usually identify them with name and country code. Cases where NSIs deliver foreign parent to EGR without proper identification can lead to creation of duplications in the system. Linking only on names is not secure; the automatic linking process cannot create a link between the existing EGR unit and the same unit provided by a NSI. This results duplication, the same legal unit is registered twice in the EGR system, the groups will be split and double as well.

The new identification service will ensure that NSIs can uniquely identify the partly known and not precisely identified foreign legal units and can get their unique identifiers and characteristics from the authentic source. The identification service will provide a 'Legal Entity Identifier' for legal units in and outside Europe. This number will be used as the primary identification number in the workflows in the EGR network. This ID should be unique and widely used. For European units the ID should be composed on the basis of a public available national identification number. Due to the possibility that two countries can use the same kind if identification system the country code as another identifier element should be added as well.

The availability of national identification number for countries outside the EU is less than inside the EU. EGR possibly should use a private source identifier to uniquely identify the foreign legal units. E.g. the DUNS number, the ID number of private data provider Dun & Bradstreet is a worldwide unique identifier used in many databases. As such it could serve as 'Legal Entity Identifier' for companies outside EU. However the use of private source identifiers is subjected to licenses. EGR should make an agreement with one private data provider to use its identification number as unique ID for countries outside the EU.

In the future also the coverage on individuals should be solved; these records usually miss from the national registers and from private sources.

To establish an efficient identification service the preferred situation is that this service keeps identifying information on all incorporated legal units included in the national statistical business registers of the NSIs. In case the identification service does not comprise the complete set of incorporated legal units, the identification service will carry out a first search in the database of a commercial data provider (if agreed with the provider). This request can consist of an individual search as well as of a list of legal units. After a successful identification the NSI of the country of the residency of the legal unit will be requested to confirm the indentifying information. For non-EU units the identification service should be developed with co-operation of one commercial data provider.

With unique identification of legal units the EGR could avoid duplications, and this will result a large improvement in quality. Foreign legal units will be uniquely identified; duplication of legal units and groups can be avoided.

3- Platform for the users

One of the main objectives of EGR V2 is enabling statisticians and register staff of NSIs and NCBs to consult, to retrieve and to update EGR data.

The main features are:

- To consult 'live' EGR data
- To update EGR data as far as allowed by the authenticity rules
- To validate EGR data

Statisticians as well as business register staff from NSIs and NCBs will be authorised to consult the 'live' EGR by an online interface and to add changes on data for which the EGR is authentic store. The provision of update rights to statisticians is considered as a critical success factor for the use of the EGR.

The needs for updated information are not necessarily the same for all the variables: the variable that needs to be validated at the earliest is the UCI (to launch OFATS surveys, for example), but the final tree structure can be provided at the end of the year, after the performers ran their surveys and when they want to impute the data.

Moreover, all the NSI will not update the EGR at the same time and will not have availabilities for validating at the same time. So, the process should be flexible enough to allow NSIs to work in their own pace. Several actors will play a role in the EGR maintenance process; four main actors can be defined:

- The statisticians: some users will have the possibility to transmit information (about UCI, about cross-border relationships, for example).
- The NSA BR staff: they should ensure that all the elements that should be checked and changed have been treated before to give his stamp.
- Eurostat central EGR: will assess the quality of the information via macro-validation procedure and send information to the NSA for micro-validation. Moreover, some information could be centrally validated.
- The profiling teams in the NSA: they should be responsible for the validation of the pre- selected large MNEs in the target of profiling.

The EGR V2 will provide the access to confidential micro data via remote access of an online interface. The development of the environment providing access to confidential data stored by Eurostat is a major challenge for the project.

With the new platform the users can consult, and propose updates to EGR data. The proposed updates - according to the authentic source principle - should be validated by EGR or NSI staff depending on the appointed authentic source. The platform will make the updating procedures transparent; it offers the opportunity to share and to communicate on new information, the status of proposed changes can be tracked.

4- Control of the quality: the evaluation/validation process

The EGR process is intended to be an all-along-year process. It means that new information will be registered at any moment, with a date of event. However, the users need pictures of the database at some points of the year. The quality of the information that is released in these pictures needs to be assessed. It is in that context that the evaluation/validation process takes place.

The next scheme gives an idea of the EGR process and its different steps.
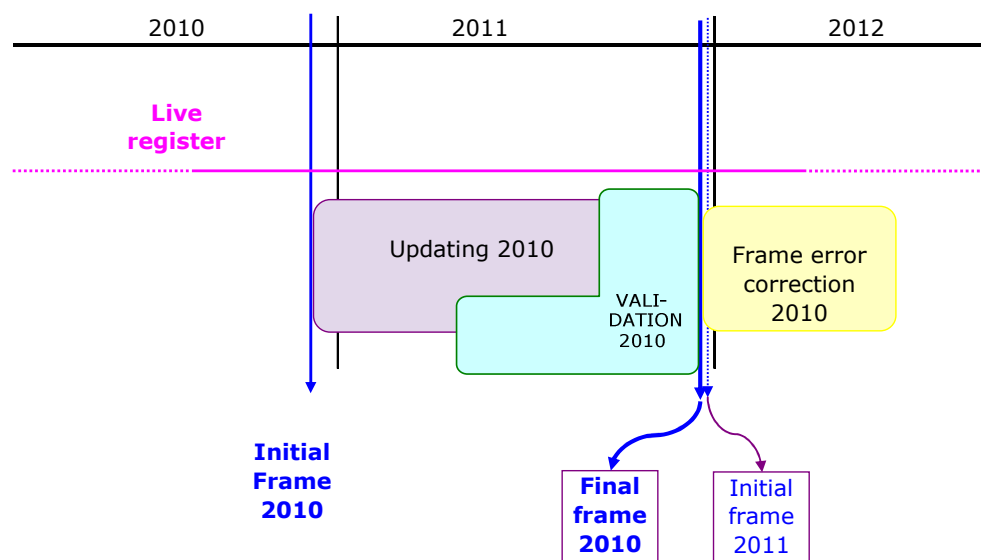
The evaluation/validation should take place after the updating of information. The preceding scheme implies that the succession of phases between updating and validation is not well delimited in the time. Indeed, the different items to be validated don't require the same treatment before to be validated.

The evaluation/validation process is considered to be handled into two parts: macro-evaluation and micro-evaluation/validation.

The macro evaluation can be defined as the "monitoring/evaluating of significant changes in the EGR frame populations". The macro evaluation should be applied to all the groups. It

should be undertaken several times until the frame is considered of enough good quality to be released. It can be seen as a set of rules that allow detecting outliers *assorted with* a decision tree that explains what to do to reduce the discrepancies according to their impact on the global figures and the size of the MNEs. This tree may define changes that are normal, groups for which a micro evaluation is mandatory and groups (or information) that can be amended automatically. It should be undertaken centrally by Eurostat.

The micro validation can be defined as the "monitoring/evaluating of changes in a significant MNE ". It can be the result of output of the macro evaluation or an independent process with separate request: this process can be defined for part of the variables or for targeted populations (for example the largest MNEs). Some external information could be used as a reference to test the quality of EGR information. From these two processes, a list of groups that the NSIs should look at should be edited centrally by Eurostat. This list could be considered as the minimum requirement, but the countries could go further in the groups' evaluation.

The variables that should be focused on have been defined. They are related to the MNE as a whole (global employment, global turnover, NACE activity code) or to parts of it (validity of the UCI, validity of the ownership tree).

PART III - The role of the users in the quality control in the EGR

The statistical users, as long as they have the closest contacts with the primary source of information (the MNEs or their affiliates) through the surveys they conduct, should have the most up-to-date information on the MNE and their components. The EGR is intended to integrate this information and to disseminate it to all the users, according to its principles (authentic source, in particular). The users could report on several points at different time of the EGR process, as soon as they get new information that is not yet in the EGR.

1- Request for adding MNEs in the EGR

The selection of the groups in the EGR has been made, up to now, according to the significance of the MNEs in terms of number of subsidiaries and geographical expansion. The EGR 2008 and 2009 covered the 5000 largest MNE operating in the EU. The 2010 population has been enlarged to 10000 and it is planned to reach a full coverage in the EGR2013. This full coverage is understood in reference with the target populations for the users, mainly FATS and FDI.

The users have then an important role to play in order to define their frame population and to check if the EGR reflects properly their needs. In a current process, they will have the possibility to ask for the inclusion of a MNE which they know being in their frame population and being absent of the EGR.

2- Intervention during the updating process

The users are often the first to know about a change in the legal structure of a MNE. During their surveys, they might have information about new relationships between two legal units, the inclusion in a MNE of a new subsidiary or, on the contrary, the removal of a subsidiary.
The MNE might report them global information which should be considered as the best information on a given MNE.

The users will be able to report every change or signal wrong information through the web platform. However, all this information need to be checked and agreed by the EGR staff in the NSI before to be integrated in the EGR (according to the authentic source principle).

The following picture takes the example of information coming from the profiling staff in the NSI. During the profiling process, the profiling team could get the more recent list of the legal units belonging to the MNE; this list is not restricted to the country of the UCI, where the profiling is preformed, but to all the countries where the MNE operates. This list could be submitted to the EGR and assess by the BR staff of all the countries involved in the given MNE.
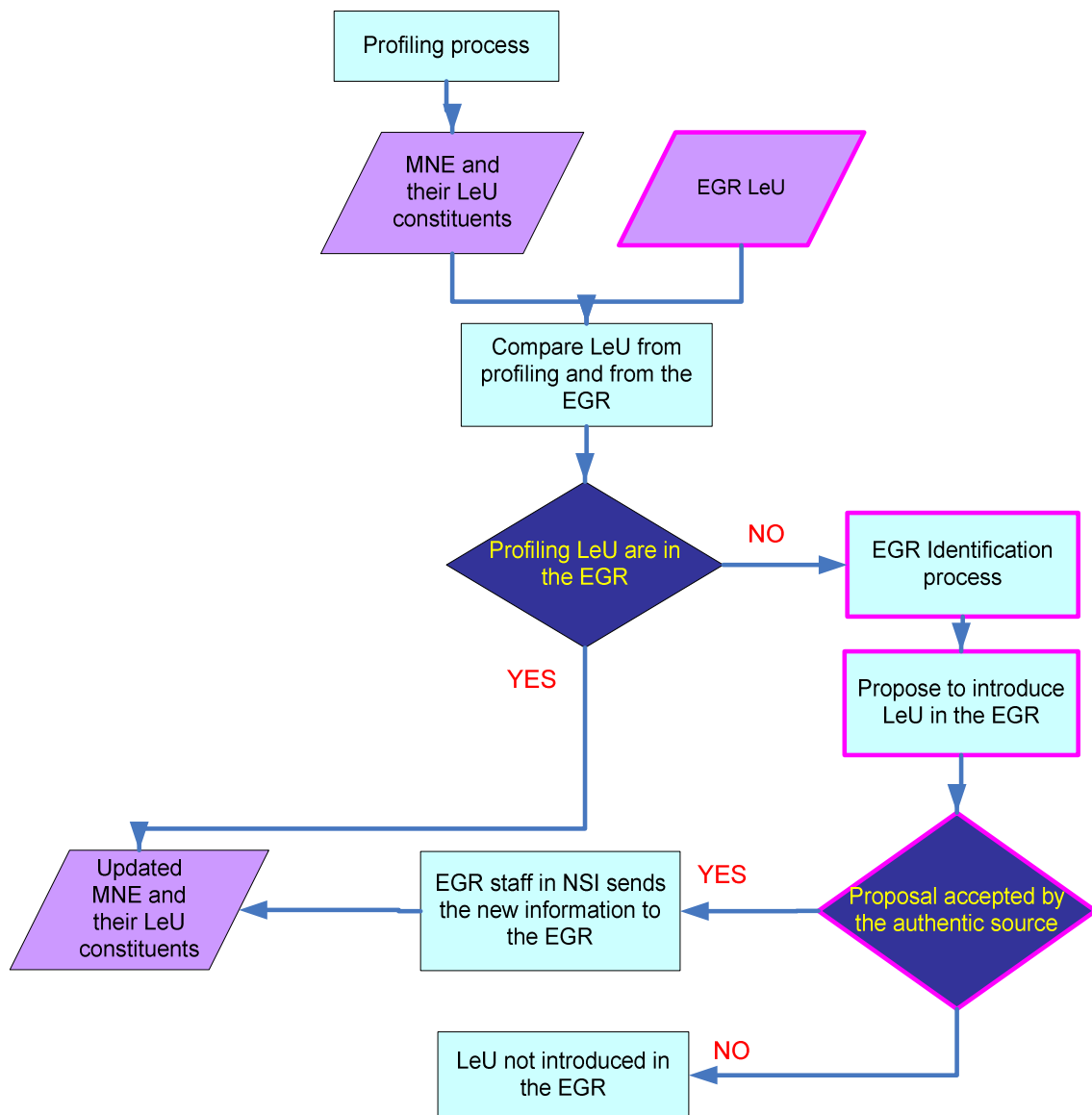
3- Request for micro evaluation

The micro evaluation/validation of a given MNE can be launched after detection of suspicious figures during the macro evaluation. It could be also made according to user information or on their request.

For example, if the users report information at the MNE level (global figures on employment or turnover, place of the UCI,…) which is conflicting with the EGR information, there will be a need to launch a validation of the all the elements concerning the MNE.

4- Some limits to the users' intervention

The EGR intends to fulfil the needs of all the statistical users involved in globalisation statistics. These users may have a different viewpoint of the MNE situation and report conflicting information to the EGR. This may cause a burden for the BR staff in the NSI if they have to treat several times the same case.

In order to avoid such situation, the EGR needs to define rules to treat conflicting information and provide, at the end, single information used by all the users.

The "single flow" principle implies that in some cases, an agreement between all the users and producers should be found at the national level. The EGR can suggest them to find a national solution that would satisfy all the parties, but this national consultation is under national prerogative.