**Quality indicators of the EuroGroups Register**

Zsolt Völfinger, Eurostat, zsolt.volfinger@ec.europa.eu

**Abstract**

*The EuroGroups Register (EGR) is a statistical business register focused on multinational enterprise groups active in Europe.*

*To create the EGR, Eurostat collects input information from the national statistical business registers and from commercial sources. After consolidation and validation, the register contains the global structure of the multinational enterprise groups for a given reference year.*

*To monitor the yearly production cycles of the EGR and to evaluate the results of the register Eurostat developed a wide range of quality indicators for the EGR process.*

*This paper presents the system of the quality indicators built behind the EGR production process:*

* *Input indicators to measure the completeness and consistency of the input data files*
* *Throughput indicators evaluating the data processing results in the different steps of the EGR production process*
* *Output indicators measuring the completeness and consistency of EGR final data*
* *Micro-level output indicators comparing EGR data to official EU statistics on foreign affiliates*
* *Development of a composite indicator to report on the overall EGR quality*

**Keywords:** business register, EuroGroups register, process quality indicators

**1. The EuroGroups Register**

The EuroGroups Register (EGR) is a statistical register of Eurostat and national statistical authorities of the European Union (EU) and European Free Trade Association (EFTA) countries.

The EGR covers the population and units of multinational enterprise groups present in the EU and EFTA countries. A multinational enterprise group is defined as an enterprise group comprising at least 2 enterprises or legal units located in different countries. If one legal unit of a multinational enterprise group is in the EU or EFTA, the group is already in the scope of the EGR.

**2. EGR production process**

The EGR produces annual data, reflecting the status of the registered enterprise groups as of 31st December of the reference year. The collection and validation of the EGR data are complex and long processes. The EGR final data for the reference years are available for users 15 months after the end of each reference year.

To create the EuroGroups Register, Eurostat collects input information on group members and on their relationships from the national statistical business registers of 28 EU and 4 EFTA countries, and from one commercial data source.

The following micro data are collected from national statistical business registers on the constituent units of the multinational enterprise groups:

* Legal units (companies): identification number, name, address details, date of birth;
* Relationships between legal units: domestic and cross-border relationships of the resident legal units, including the identification number of the subsidiary and the parent, type of the relationship (control or minority), percentage of the ownership and control;
* Enterprises of the delivered resident legal units (statistical representation of companies): identification number, name, address details, date of birth and economic characteristics as NACE main activity code, number of persons employed, turnover.

The EGR also acquires data from one commercial data provider. This acquisition ensures coverage for units outside the EU+EFTA and consistency of the largest enterprise groups.

After validation of input data from the different sources the EGR calculates the consolidated picture on the legal units, on their relationships and enterprises.

Complementary and conflicting information from the sources are treated in the EGR process with predefined priority settings and business rules. In the EGR production the highest priority data are the data from the country of the subsidiaries (bottom-up view), followed by the data from the country of the parent companies (top-down view), followed and complemented by data from commercial data provider.

Having final legal units and relationships for a reference year the EGR calculates the enterprise group structures. The enterprise groups are built on those legal units, which are linked together by control relationships, where the voting rights are above 50%.

The following enterprise group characteristics are calculated in the EGR process for each group: EGR identification number, group main activity code, group employment, group turnover and country of global decision centre (country of the group).

Before publishing the final data on the groups, the calculated group structures and group characteristics are revised and validated by the participating countries and Eurostat.

The EGR process was redesigned to EGR 2.0 from 2012, the new process was fully implemented in 2015. The switch to EGR 2.0 improved both the EGR process and the IT system. Data for reference years 2014, 2015 and 2016 were produced completely with the EGR 2.0 system.

EGR 2.0 aims to cover all relevant multinational enterprise groups active in the EU. For reference year 2014 EGR produced data of 61 000 multinational enterprise groups, for 2015 data of 80 000 groups. For the last available reference year, 2016 EGR produced data on 111 000 groups covering 779 000 enterprises.

**3. Quality indicators in the EGR process**

EGR 2.0 processes every year big volume of data from multiple sources. In order to facilitate and to monitor the data processing, three types of quality indicators were defined for the EGR 2.0 system: input indicators, throughput indicators (process indicators) and output indicators.

The input quality indicators measure the completeness and consistency of the national input files, the throughput indicators measure the changes of source data during the EGR data processing, the output quality indicators are calculated for measuring the completeness of the EGR output data which are disseminated for users.

*3.1. Input data and their indicators*

The EGR input quality indicators measure the completeness and consistency of the national input files (legal units, relationships, enterprises, enterprise groups). All input indicators are calculated by the EGR system based on the national input files data.

Input indicators calculated in the EGR:

* Number of distinct legal units received for the reference year, for the previous year and their ratio.
* How key indicators on foreign ownership or foreign control are filled in the national files on legal units.
* Number of distinct relationships received for the reference year, for the previous year and their ratio (see Table 1 as example).
* Number of enterprises received for the reference year, for the previous year and their ratio.
* How main variables (employment, turnover, main activity code) are filled in the national files on enterprises (see Table 2 as example).

Eurostat monitors the input indicators during the process and contacts the contributing national statistical institutes (NSIs) where issues are detected, for example the number of received records significantly decreases between two years or where the key variables of the input files (e.g. employment) are not sufficiently populated.

The monitoring of the indicators results more complete input data for EGR, e.g. for EU enterprises in 2015 the main activity code was filled for 91% in 2016 for 99%.

**Table 1: Input quality indicators on EGR relationship data**

Source: EuroGroups Register 2016 data

Notes: Columns of the table: source of data, reference year (T), number of all relationships received for year T, number of all relationships for year T-1, % T/T-1, number of cross-border relationships received for year T, number of cross-border relationships for year T-1, % T/T-1

**Table 2: Input quality indicators on EGR enterprise data**



Source: EuroGroups Register 2016 data

Notes: Columns of the table: source of data, reference year (T), number of enterprises received for year T, number of enterprises for year T-1, % T/T-1, % of records with NACE main activity code, % of records with employment data, % of records with turnover data

Eurostat and the contributing NSIs are working on defining thresholds as minimum criteria for acceptance of input data. Below these thresholds national data will be declared as of insufficient quality and new file delivery will be requested from the NSIs.

*3.2. Indicators on data in the EGR production*

The throughput indicators measure the changes of source data during the EGR data processing. Data on legal units, relationships, enterprises and enterprise groups are monitored.

These indicators follow the data in the EGR system from the source area, where the input data are loaded, via the transformation area, where the input data are standardised and transformed until the consolidation area with final data on EGR units.

In the EGR process the data of the units are stored in three different statuses:

* The input data of EGR are stored in the input area of the database, as they were delivered by the sources.
* In the transformation area the input data are transformed and standardised, all records are identified, all values and codes are transformed according to the EGR standards. The transformation area is the location where the standardised records wait for further processing in the consolidation.
* In the consolidation process all the data from the different data providers are combined to create final entities for the reference year. The consolidation area is the final view on EGR units after processing, where the only information from the highest ranked source is stored on the units.

The throughput indicators measure the proportion of the records which reach the consecutive steps of the EGR process, ideally all, or in some cases most of them should reach the consolidation area. All throughput indicators are calculated by the EGR system.

The following throughput indicators are calculated in the process:

* Legal units in the source area; transformation area; consolidation area and difference of consolidated legal units and transformed legal units.
* Legal units in the consolidation area in the reference year; in the previous year and the difference.
* For relationships similar indicators are calculated, with a split between all and resident relationships (see Table 3 as example).
* For enterprises the quality indicators measure the number of enterprises in the different stages of the EGR process.
* For enterprise groups the indicators monitor the number of groups between two reference years and the cases where continuity exists between the two years (see Table 4 as example).

Eurostat monitors the throughput indicators during the process and reacts on eventual problems in the process flow. Where needed, the concerned NSIs are contacted by Eurostat during the process.

**Table 3: Throughput quality indicators on EGR relationship data**



Source: EuroGroups Register 2016 data

Notes: Columns of the table: source of data, reference year, number of resident relationships received and imported, number of resident relationships transformed, number of resident relationships consolidated, difference of consolidated and imported resident relationships, % of consolidated and imported resident relationships (records reaching the consolidation area)

**Table 4: Throughput quality indicators on EGR enterprise group data**



Source: EuroGroups Register 2016 data

Notes: Columns of the table: country of the group, reference year (T), number of enterprise groups in year T, number of enterprise groups in year T-1, difference of year T and T-1, % T/T-1, groups with continuity, % of groups with continuity

Also for the throughput indicators Eurostat and the NSIs are working on defining thresholds as minimum criteria for acceptance of process results in EGR. Below these thresholds new input data should be collected to meet the process requirements.

*3.3 EGR output data and their quality indicators*

In the EGR database the output indicators measure the completeness of the EGR output data, as they are disseminated to users after closing the EGR yearly cycles. Eurostat calculates these indicators based on the yearly EGR final data.

The indicators are calculated per country and measure the completeness of the legal unit, enterprise (see Table 5 as example) and enterprise group data in EGR final frame.

**Table 5*:* Output quality indicators on EGR enterprise data**



Source: EuroGroups Register 2016 data

Notes: Columns of the table: country of the enterprise, reference year, % of enterprises with: EGR ID number, national ID number, name, address, city name, country code, NACE main activity code, employment data, turnover data

**4. Measuring EGR quality compared to other official statistics**

After completion of the EGR cycles the contributing NSIs calculate micro level quality indicators on the EGR output for comparing national EGR data and national populations of Foreign Affiliates Statistics (FATS).

The objective of these EGR-FATS indicators is to test the quality of the EGR as a statistical frame population with respect to FATS. Completeness and accuracy of the two frames are measured.

The indicators are calculated by NSIs with record-level comparison of the national EGR and FATS populations, on one hand to the population of resident foreign-controlled enterprises (national IFATS population), on the other hand to the resident units on top of multinational groups (national OFATS population).

Four indicators are calculated for the EGR frames:

* Completeness of EGR - completeness of EGR enterprises compared to the enterprises covered by IFATS.
* Completeness of EGR on employment - completeness of EGR on employment of the enterprises compared to the employment of enterprises covered by IFATS.
* Accuracy of the UCI - accuracy of the group of units present both in EGR and in IFATS populations (whether they are members of the right group or not).
* Completeness of resident groups in EGR - comparison of EGR's resident groups to national OFATS population.

Table 6 presents the results of the EGR-FATS quality indicators for reference year 2015.

**Table 6: Results of the EGR-FATS quality indicators for EGR 2015 data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Completeness of EGR** | **Completeness of EGR on employment** | **Accuracy of the UCI** | **Completeness of resident groups in EGR** |
| 0.59 | 0.84 | 0.82 | 0.45 |

Source: EuroGroups Register 2015 data and national FATS 2015 data

In Q4 2018 the above listed quality indicators will be calculated for 2016 EGR data. Based on the 2015 and 2016 results NSIs and Eurostat will analyse, what are the differences in the national populations, what is missing from EGR and how can they be added to later EGR frames.

**5. Composite indicator on the EGR quality**

Mainly for project management purposes but also for simplified visualisation of the EGR quality Eurostat started to develop a composite indicator on EGR quality.

Eurostat decided to use the four EGR-FATS quantitative indicators for an overall quality measurement: completeness of EGR on units and on employment, on accuracy of UCI and on completeness resident groups in EGR.

The composite EGR quality indicator is the average of the four available micro-level EGR-FATS indicators. Non-quantitative quality criteria like timeliness of the final frame, accessibility, coherence and clarity will be attached to the composite quality indicator in a written evaluation.

The average of the four micro-level indicators resulted **68% as overall quality indicator for the EGR 2015 data**. These indicators will be calculated also for the EGR 2016 cycle and EGR progress can be monitored also with this composite number.

**6. Summary and conclusions**

Eurostat and the participating national statistical authorities are continuously working on the improvement of the EGR 2.0 quality.

With the support of the EGR quality indicators the actors can measure in the system the quality of their contribution and the quality of the interim and end results of the process.