

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

Brussels, 21-23 September 2015

Zsolt VÖLFINGER
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 identification service

Introduction

This paper will present the solution, how the unique identification was implemented in the EuroGroups Register. The created identification system will be explained and the applied unique identification number will be presented.

1. Identification of legal units in business registers

1.1. Identification of units in national business registers

Statistical business registers cover the resident business populations registered in the countries. Based on administrative or survey data the identification of the resident legal units in statistical business registers is a relatively simple task for statistical offices.

For the mapping of multinational enterprise groups the registration of the cross-border relationships is necessary. Cross-border relationships include foreign units; these foreign entities also should be registered in the national statistical business registers or linked to it.

In the European Union (EU) and European Free Trade Association (EFTA) countries the registration of parents or subsidiary companies is compulsory in statistical business registers as above 10% relationships. The relevant European Regulation (EC) No 177/2008 requires EU member states to register the ownership of units if these data are available from administrative data sources.

For those relationships where the parent or the subsidiary of a legal unit is resident of a foreign country the national administrative and survey sources usually deliver limited information on the foreign entities, normally only the name and the country of the foreign entity is registered without registration of a unique identification number.

1.2. The EuroGroups Register – process and identification

The EuroGroups Register (EGR) is the statistical business register of multinational enterprise groups of the EU member states and in the EFTA countries.

In order to create the EGR, Eurostat collects enterprise group information from the national statistical business registers of the EU member states and participating EFTA countries and commercial sources. All sources report to the EGR the details of the constituent legal units of the groups and the relationships between the legal units. After the consolidation and validation process based on the pairwise relationships the EGR creates the global structure of the groups.

National register staff and statistics compilers receive access to all units of the multinational enterprise groups, if at least one unit of the group is located in their national territory. These populations can be used for survey statistical survey frames at national level.

The EGR process is in operation since 2009. In the first cycles of the EGR the identification of the foreign entities caused several problems and created duplicates. The EGR system was linking the legal units on national ID numbers, as they mostly were not available, on their simplified names. As the EGR has 32 national sources and 2 commercial sources the linking was very difficult and resulted several duplicates in the EGR system.

1.3 Ongoing efforts for creating common identification numbers

In the past years several projects were started in the EU or globally with EU participation with the goal to create unique identifications numbers for certain group of legal units.

The Global Legal Entity Identifier System (GLEIS) is an initiative launched in 2011 by the Financial Stability Board, mandated by the G-20 (20 major economies, 19 individual countries along with the EU). The purpose is to facilitate coordination amongst the regulatory community through the creation of a global Legal Entity Identifier (LEI). The LEI is a 20-digit code to uniquely identify (worldwide) legally distinct entities that engage in financial transactions. The GLEIS aims at covering all legal entities worldwide. Currently it covers all legal entities involved in financial transactions.

The Register of Institutions and Affiliates Database (RIAD) was launched by the European Central Bank and accessible to all members of the European System of Central Banks (ESCB). It holds reference data of mainly financial organisational units relevant for ESCB's statistical departments and other business areas, predominantly banking supervision and market operations. The main focus of RIAD is on financial institutions. Units outside the financial sector are currently only covered on basis ad hoc needs of users.

Because of restrictions and limitations of the ongoing EU or global projects the EGR project decided to create an own identification number within an own identification system. On the other hand the data structure of the EGR will be able to apply any time one of the future global identification numbers.

In the redesign of the EGR system since 2012 to EGR 2.0 one of the main goals was to create a system assigning unique identification numbers to the legal entities applied in the data exchange between member states. A new tool was designed and developed by 2013; this was the EuroGroups Register identification service (EGR IS).

2. The EuroGroups register identification service

The EGR IS is an identification system with statistical purposes, launched by Eurostat and the EU and EFTA member states' statistical offices. The EGR IS is a module of the EGR system. The EGR IS is an application assigning the unique identifications number for legal units in the EGR process. It is an on-line application with central database and with remote access for users in the EU and EFTA countries.

It allows users to identify legal units and assigns unique identification numbers to these units. The EGR IS aims to cover all incorporated legal units that may potentially be part of multinational enterprise groups. Currently the EGR IS covers 16 million legal units, of which 15.5 million from EU and EFTA countries.

The users can connect to the user-interface of the EGR IS for identification of the foreign legal units they store in their national statistical business registers. The unique identification of legal units is critical for avoiding duplications in the EGR. Proper identification of foreign legal units is an important step towards achieving interoperability of the national business registers.

Any national statistical office involved in exchanging data with the EGR must use the unique identification numbers when communicating with the EGR. For instance, when member states' national statistical authorities inform the EGR of relationships between resident and foreign legal units, both units should be identified with the ID numbers assigned by the EGR IS. For the identification the EGR IS is applying an own identification number, called the legal entity identifier number (LEID number).

3. The EGR legal entity identifier number

The legal entity identifier number, LEID number is the unique identification number assigned by the EGR IS. The structure of LEID numbers has been created by Eurostat specifically for the purposes of the EGR; an LEID number is a combination of the country code, register code and national identification number of the legal units.

3.1. Format of the LEID number

For EU and EFTA units the LEID number is composed of the ISO 2-digit country code, the internal register code assigned to the national register by the EGR (referred to as the national identification system code or NIS code) and the legal unit's national identification (national ID) number, as assigned by this same national register.

The types of national ID number used in the EGR were chosen by the member states' national statistical offices. The aim was to select types of ID numbers that are also widely available from commercial sources in order to simplify the matching to be carried out in the EGR process. In addition several member states specified secondary national ID numbers that are stored in additional national registers. They are not used for the creation of the LEID number but can be important for linking the records held by the EGR IS to the records acquired from commercial data sources.

For extra-EU and EFTA units the LEID number has the format: ISO 2-digit country code and an 11-digit sequential number.

3.2. Examples for the LEID number

The LEID number generated by the EGR for legal units based in Austria is a combination of the ISO 2-digit country code 'AT', the Austrian trade register's internal register code '21013' and the identification number given to a particular legal unit within this register, e.g. '094342p'. The LEID number for this legal unit would therefore be 'AT21013094342p'.

The LEID number generated by the EGR for legal units based in Belgium is a combination of the country code 'BE', the Belgian business register's internal register code '21021' and the identification number given to a particular legal unit within this register, e.g. '0604702156'. The LEID number for this legal unit would therefore be 'BE210210604702156'.

Annex 1 of this document presents the LEID numbers of legal units in EU and EFTA countries.

4. Sources and updating of the EGR IS

4.1. Business registers held by the national statistical authorities

The EGR IS uses a database for searching, matching and identifying the legal units. The database is maintained with data of the national statistical authorities of the EU and EFTA countries and with data from commercial data providers. The role of the national statistical authorities in maintaining the EGR IS database can vary:

- *Authentic store national statistical authorities*

National statistical authorities which send data on the full population of incorporated legal units to the EGR IS are called authentic store. The register of the national population of incorporated legal units is updated once a year. Where the national statistical authority is classified as an authentic store, it has the exclusive right to register new national legal units with the EGR IS. Commercial data providers cannot request new legal units to be added on the basis of their data. In 2015 out of the 32 EU and EFTA countries 26 countries are authentic store for resident legal units.

- *Partially authentic store national statistical authorities*

National statistical authorities that provide the EGR IS with details of only those incorporated legal units which they know belong to global enterprise groups are called partially authentic store. The

register of the national population of incorporated legal units is updated once a year. Where the national statistical authority is classified as a partially authentic store, both it and commercial data providers have the right to register new legal units in the database. 6 countries are partially authentic store for resident legal units in 2015.

4.2. Commercial data providers

Where the national statistical authority of an EU or EFTA country is a partially authentic store, details of new legal units can be sent to the EGR IS by both the national statistical authority and by commercial data providers. Information already received from the national statistical authority in relation to a particular legal unit is not overwritten by information on that same unit provided by a commercial data provider. Legal units in EU and EFTA countries can be added to the database on the request of commercial data providers on the condition that the legal units have a valid national ID number.

Commercial data providers are the only source of information on legal units outside the EU and EFTA countries, as the national statistical authorities' business registers only contain information on domestic legal units. Legal units from outside the EU and EFTA countries can be added to the database without a national ID number, i.e. they are not subject to the requirement mentioned above for legal units based in EU and EFTA countries. In the large majority of cases in fact, the national ID number is missing when these units are registered, and the LEID number is therefore composed of the country code and an 11-digit sequential number.

Commercial data providers provide the EGR process with data on legal units and their relationships. The information sent by the commercial data providers is matched up to the records held by the EGR IS. For those units for which records already exist, these are updated with the new information; for those units for which no record yet exists, a new record is added to the system. In addition to the fact that commercial data providers send data to the EGR on a regular basis, users of the EGR IS database can also search for legal units in the databases of these commercial sources, via a built-in web service within the EGR IS. They can then add commercial records for legal units from outside the EU and EFTA countries to the database at any time.

5. Functionalities of the EGR IS

The EGR IS provides a means of retrieving identification information relating to legal units. Registered users can search via the online interface of the EGR IS.

In addition to the LEID number, the EGR IS also contains other information serving to identify a particular legal unit, e.g. the name and address of the unit and/or additional ID numbers. Users can search for legal units using any of these pieces of 'identifying information' and can then download the full file of information relating to that unit into their national statistical business registers.

5.1. Manual identification of legal units

The online interface of the EGR IS allows users to search for one specific legal unit. To carry out a manual search, the user must enter the country in which the legal unit is based, and can then add additional search criteria such as the company name, the city in which it is based, its address, its national ID number, commercial ID number(s) or other ID number(s) (e.g. VAT number). At least one of the company name, the national ID number and the commercial ID number should be entered when performing a search.

The EGR IS provides the user with the results of their search immediately on screen. Where there is just one single result, e.g. for a search on a national or commercial ID number, the identification information of the legal unit that matches the search criteria entered will be displayed in full. Where the search produces multiple results, e.g. for a search based on a name, up to 20 possible matches will be displayed by the system, sorted in descending order on the basis of a score indicating the similarity. The user can select any one result from the list to see the full identification information relating to this legal unit.

Image 1 – Manual identification screen from the EGR Identification Service

logged in as volfizs (NSA_NA_HU_BR) | LOGOUT

EuroGroups Register
Identification Service

European Commission > EGR Identification Service

Online identification | Batch identification | Datasets management

Identifying data

Country: --choose a country-- | Name: | National ID: | DUNS: | BVD ID: | LEID: |

▶ Hide advanced search:

NSA ID: | EORI: | LEI: | VAT: |

Address: | Postal code: | City: |

Run identification

Data sources

If the EGR IS does not find any results for a search relating to a legal unit from outside the EU and EFTA countries or a legal unit based in a country whose national statistical authority is a partially authentic store, the user can resubmit the request and the system will then search for the legal unit within the database of a commercial source.

Once the request has been resubmitted, the EGR IS will communicate with the database of a commercial data provider via a web service. The user does not have to leave the EGR IS in order to do this, as this application will forward the request. The commercial data provider searches its database for matches, based on the country code and the name of the legal unit, and the three closest matches are then sent to the EGR IS. The user can select from these results the one that best matches their search criteria and the details of this legal unit will then be obtained from the commercial source and added to the EGR IS database. In a later search other users will be able to find this legal unit within the EGR IS, and the same request will not have to be resubmitted to the commercial data source.

If the EGR IS does not find any results for a search relating to legal units based in countries whose national statistical authorities are authentic stores, the user should refine their search criteria and perform a new manual search. This type of search cannot be resubmitted to the commercial data provider because new legal units cannot be added for authentic store countries to the database on the basis of information from commercial data providers.

5.2. Batch identification of legal units

Where users need to identify a set of legal units, they can send files containing the search criteria to Eurostat via the eDAMIS secure communication platform. The minimum required search criteria when using this method are the country code and the name of the legal unit. The user can also include any other available information (e.g. the city in which the unit is based, its address, its national ID number, commercial ID numbers or other ID numbers), in order to make the search more efficient.

Eurostat uploads the identification files onto the EGR IS and the user can see and manage their data files via the system's interface. The user can then launch the batch identification search and the matching results will appear directly on screen. The searched legal units will be classified to 3 groups:

- For searches which can produce only a single result, e.g. a search based on a national or commercial ID number, the records with match will be classified as 'identified'.

- For searches where there can be multiple results, e.g. for a search based on a name, the records with match or matches will be classified as *'open for identification'*. The open records can be analyzed by the users case by case on the online interface. Up to 20 potential results will be displayed by the system, sorted in descending order on the basis of how well they match the search criteria. The user can then select any one from the list as best match. Searches for legal units located outside EU and EFTA countries and in countries whose national statistical offices are partially authentic stores can be resubmitted to a commercial source. The search request can be resubmitted for all the legal units to which the request relates in one step. The process is similar as for single manual identifications.
- If the matching process carried out by the EGR IS does not produce any results for a legal unit based in a country whose national statistical authority is an authentic store, the record will be classified as *'not identified'*. In such cases, the user will have to perform a new search for the same record using different data, e.g. entering an alternative name.

Image 2 – Batch identification screen from the EGR Identification Service

logged in as volfziz (NSA_NA_HU_BR) | LOGOUT



EuroGroups Register

Identification Service

European Commission > EGR Identification Service

Online identification
Batch identification
Datasets management

Identification request number: 592

Basic info		Legal unit info	
Dataset name	EGR_ISNORLE_N_LT_2014_0001_V0001_060820 ↕	Number of Legal Units	3409
Request date	2015-08-17	Resident Legal Units	0
Request status	FOR_VERIFICATION	Non resident EU Legal Units	2705
		Non resident non EU Legal Units	704

Identification process (After source with priority: 1)

Identified	Open for identification	Not identified
Number of Legal Units: 1910	Number of Legal Units: 1350	Number of Legal Units: 149
Resident Legal Units: 0	Resident Legal Units: 0	Resident Legal Units: 0
Non resident EU Legal Units: 1781	Non resident EU Legal Units: 775	Non resident EU Legal Units: 149
Non resident non EU Legal Units: 129	Non resident non EU Legal Units: 575	Non resident non EU Legal Units: 0

Identified
Open for identification
Not identified

Identifying data						
COUNTRY	CITY	NATIONAL ID	DUNS ID	BVD ID	NAME	ACTION
FI		1917540-9			ELTEL GROUP CORPORATION	i
CY		10120311			ELANTIS INVESTMENTS LTD.	i
FI		1944757-4			ALMA MEDIA OYJ	i
DK		19475344			DAN-LAT AGRO APS	i
FI		1948808-5			EPI BALTIC OY	i

Items per page: 5
Page: 1 / 382

Close request
Resubmit
Show source
Back

EGR 5.0.5 | [Top](#)

The user can close the request for batch identification at any time during the process. Ideally this would be when all the results have been confirmed - either following the search in the EGR IS database, or after having resubmitted the records for a search to be performed in a commercial source database. Once the identification request has been closed, the EGR IS will generate an output file containing the identification information relating to the confirmed matches. The records that were not able to be identified will be also displayed in the output file. Eurostat will send the output file to the user via the eDAMIS secure communication platform.

5.3. Management of datasets

Users can manage their datasets on a separate page of the EGR IS application. The datasets can include resident legal units to update the resident population of legal units in the EGR IS and non-resident legal units that the user is wishing to identify. The EGR administrators at Eurostat are able to manage the commercial data providers' data files via this page, in order to update the EGR IS.

6. Search mechanism applied in the identification process

Where searches are performed using ID numbers, e.g. national or commercial ID numbers, the EGR IS will produce either one single result or no results. Where a search has multiple results, e.g. a search based on a name, the system may display up to 20 possible matches, sorted in descending order, on the basis of how well they match the search criteria. It is only when a search is performed using text, that the system computes the closeness of each match.

The EGR IS computes the closeness of the match on the basis of the Levenshtein distance algorithm. For each of the criteria provided, the distance between the criterion value (i.e. the term entered in the search) and the entry in the corresponding field of the record suggested as a possible match is computed. A weighted average is then taken of the 'similarity score' for each criterion. This is shown as this result's 'final score' on the search results panel.

The criteria used in the full text search have the following weights: name of the legal unit: 4; country code: 1 (the country code must match exactly); address: 1; postcode: 1; city: 1. Where the data for the address, postcode or city is missing or has a low 'similarity score', the system will not take these fields into account in calculating the weighted average.

7. Access to the EGR IS

Access to the EGR IS is restricted to users at national statistical offices of the EU and EFTA countries. Users need to have an account and a password in order to access the system. The system is having close to 200 registered users in 2015.

8. Identification results of the EGR IS

The national statistical offices of the EU and EFTA countries use the EGR IS since 2014. Substantial progress was reached in the past time in the identification of foreign legal units from the national registers despite the heavy initial workload.

Normally the batch identification cannot deliver 'identified' status for all foreign units, for such results a match on ID numbers or high similarity if names are requested. The results of the automated matching are dependent on the quality of national data on foreign entities (e.g. quality of company names, availability of foreign ID numbers).

For the rest of the cases, where the automated matching does not deliver one-to-one identification, case by case manual identifications actions should be done by member states.

The labor-intensive manual matching process is ongoing since 2014 in the member states. Eurostat and the member states are targeting to identify the large majority of the foreign legal units stored in

national business registers by the end of 2015. Status reports will be sent to Eurostat by the end of 2015.

The goal of Eurostat and the member states is to identify all foreign units in the national business registers in the EGR IS and continuously work on the maintenance of the population of foreign legal units.

Annex 1 – The structure of LEID numbers of legal units in EU and EFTA countries

Country	Country code	Register code	National ID format
Austria	AT	21013	094185h
Belgium	BE	21021	207484681
Bulgaria	BG	21023	175375983
Croatia	HR	22097	60698900286
Cyprus	CY	25054	10011130
Czech Republic	CZ	21055	49551604
Denmark	DK	21058	30179269
Estonia	EE	21063	11060531
Finland	FI	21069	2470098-4
France	FR	21074	434385290
Germany	DE	21056	20355HRA112456
Greece	GR	22088	95283423
Hungary	HU	22099	40689928
Iceland	IS	21108	4608992229
Ireland	IE	21101	265676
Italy	IT	25109	018554305
Latvia	LV	21134	40002154526
Liechtenstein	LI	21128	FL00022987125
Lithuania	LT	21132	110797780
Luxembourg	LU	21133	B159712
Malta	MT	21152	C51547
Netherlands	NL	31165	34343296
Norway	NO	23166	871582602
Poland	PL	21178	145875000
Portugal	PT	22183	500780927
Romania	RO	22188	18858423
Slovakia	SK	25201	31389490
Slovenia	SI	23199	5858801000
Spain	ES	22067	B62843974
Sweden	SE	23196	5020391131
Switzerland	CH	23043	E107534059
United Kingdom	GB	21076	8356904