

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

Rico Konen
Statistics Netherlands

Session 6: Profiling and Globalization

Collaboration SBR and EGR

Abstract

This paper describes the collaboration between the Dutch SBR and the different interfaces of the EGR: EGR-IS, EGR-IM, EGR-CORE, EGR-FATS. But also the complementary actions executed by the FATS statisticians to improve the quality provided by the EGR, regarding the allocation of the GDC, are discussed in this paper.

Since the start of the EGR back in 2007, Statistics Netherland (SN) was deeply involved in the methodology and the development of the different applications of the EGR. However the processes to exchange data with the EGR, to use the EGR applications itself or to distribute the data extracted from it, had not yet been optimized.

Other topics for improvement within SN were related to the identification of cross border relationships and the exchange of data with the EGR. Since 2014, information about approximately 20.000 cross border relations were provided by the Dutch Chambers of Commerce. The identification process and the maintenance using the EGR-IS, meant a big increase of the workload of SBR staff. Besides, the statistical processes to exchange data with the EGR, depended strongly on the availability and the efforts of some specific staff members of the Dutch business register department. This might have affected not only the quality of the information provided by SN in the historical EGR cycles, but might also be a risk factor for future deliveries.

In order to overcome the drawbacks, SN developed in 2017 a special Cross Border Register (CBR) and a special EGR-satellite. The EGR satellite is based on the coordinated data extracted from the final frame(s) in the EGR-FATS application. The aim of the satellite is to provide a coordinated frame to be used for global statistics about the Netherlands in general. The aim of the CBR is to implement the EGR methodology, facilitate the data exchange with the different EGR applications and support the identification process in order to coordinate and ease the workload of the SBR staff.

At this moment, the quality of the information provided by the EGR for the Netherlands is not yet sufficient to produce Dutch FATS statistics. FATS statisticians still discover inconsistencies in the allocation of the country of the GDC by using different information sources. Sharing some information in this paper about the sources and methodology they used to detect these anomalies, might contribute to raise the quality of EGR data.

Keywords: EuroGroups Register (EGR), Statistical Business Register (SBR), Cross-Border Register (CBR), Global Decision Centre (GDC), Foreign Affiliates Trade Statistics (FATS)

1. Introduction

Global statistics become more and more important for Europe to compare the economy of the European Union with the economies of other continents, to guard the traffic of goods and services within or outside its borders and to development policy to improve or support European interests in a global environment.

In this respect, the FATS statistics contribute to provide cross-border indicators about the influence a specific country has in the national economy of another country. The data collection process however, is complex because each organization involved in a cross border relationship has its own priorities, aims and agreements. FATS statistics are useful to explain how businesses are distributed internationally. These statistics fall apart into the Inward (IFATS) and the Outward FATS statistics (OFATS). The IFATS statistics provide an answer on how much an economy of one 'domestic' country is influenced by other 'foreign' countries, because a global decision centre (GDC) is hosted by another country than the country of the production unit. The Outward FATS statistics answers a similar question from the point of view on the country that is hosting a GDC. It provides an answer on much influence a 'domestic' country has on the economies of other countries, because the GDC controls one of more subsidiaries in the other countries.

Eurostat supports the statisticians involved in the production of the FATS statistics by developing the EuroGroups Register (EGR) in order to achieve more coherence in statistics and a better harmonization of the national results. The EGR is a part of the project ESBRS, which links all SBRs of each National Statistical Institute (NSI) and have those more harmonized.

The EGR is a system created for statistical purposes only, and to improve the quality of information on multinational groups in the European Statistical System (ESS). The EGR is built around a Live Register where microdata about legal and statistical units is used as input to produce coordinated frames about Multinational Enterprises or what in the EGR is called the Global Enterprise Group (GEG). For each GEG some characteristics, the (national) enterprises and the structure in terms of controlled legal units registered in at least 2 different countries is stored and maintained for one specific reference year (see information model in the [Annex](#)). A frame can be extracted from the EGR and used by European statistical authorities responsible for producing official statistics, national statistical institutes and authorized national central banks.

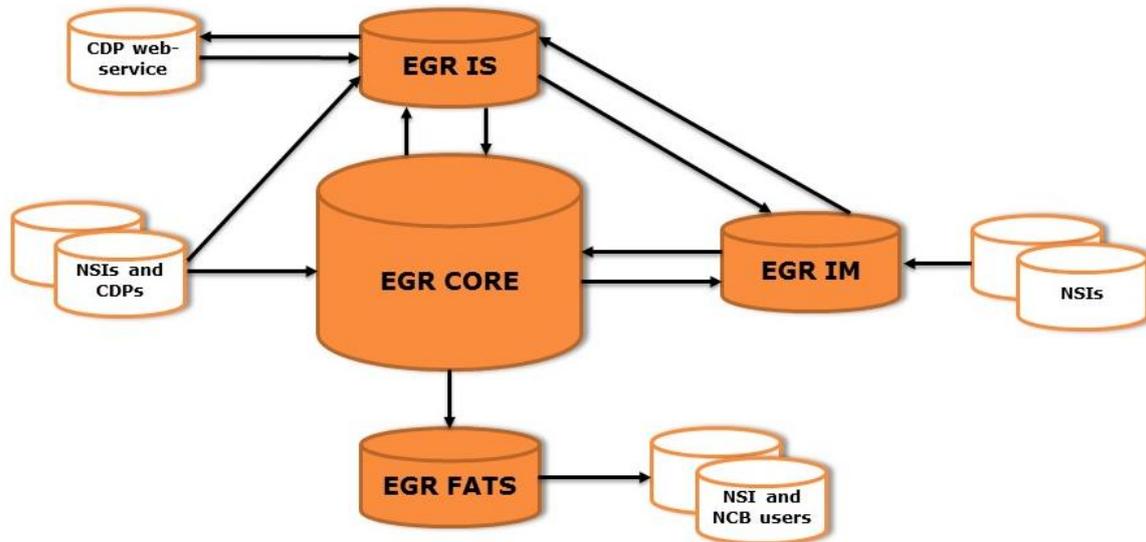
The development of the present version of the EGR (2.0) is the result of a lot of methodological and technical research by business register staff and FATS statisticians. Eurostat launched a wiki page on the EGR¹ which provides information on

- History about the development of the EGR
- Aim and mission of the EGR
- Methodology and concepts applied in the EGR
- Procedures and governance around the EGR
- Applications of the EGR

¹ webgate.ec.europa.eu/fpfis/wikis/display/EGR

The EGR 2.0 process is an annual production process which involves all NSIs of the European members (incl. EFTA countries), two commercial data providers Dunn & Bradstreet (D&B) and Bureau Van Dijk (BvD) and some National Central Banks (NCB). The processes of the EGR are based on the data collection on legal units, relationships and enterprises. The EGR consolidates the input information and create frames which consists of Global Enterprise Groups (GEG) composed by the control-relationships between legal units registered in different countries. The exchange of microdata with the EGR is supported by eDAMIS and the EGR applications as listed below.

Figure 1: Actors and applications of the EGR



Each EGR application is independent of another EGR application and can be used by multiple users. The heart of the EGR is the EGR Core application. This is the only application which is not accessible by the statisticians from a NSI, but is restricted to the users of Eurostat.

EGR Identification Service (EGR IS) ²	Application supporting the EGR and its users in identifying legal units globally by a Legal Entity Identifier (LEID).
EGR Core application (EGR CORE)	The 'heart' of EGR 2.0 system, it stores, transforms and consolidates input data of different sources and generates the EGR frames.
EGR Interactive Module (EGR IM) ³	Interactive web interface to browse, validate and update data in the live area of EGR CORE (i.e. consolidation area).
EGR FATS online interface (EGR FATS) ⁴	Application providing users a web interface to browse and download coordinated EGR frames (see information model in the Annex).

² <https://webgate.ec.europa.eu/egr-is>

³ <https://webgate.ec.europa.eu/egr-im>

⁴ <https://webgate.ec.europa.eu/egr>

The present version of the EGR processes all required information to produce a FATS master frame for one reference year. This master frame comprises the characteristics of all GEGs which have subsidiaries in at least 2 different member states. It also comprises information about the perimeters of the GEG. The perimeter is presented by the subsidiaries of the GDC. In other words the identification of the GDC and the identification of each legal unit registered in a different country of the EU it controls. To achieve this, it was necessary to introduce a common Legal Entity Identifier (LEID) to identify legal entities in the EGR for each country.

The steps of the EGR yearly cycle of reference year T is as follows:

Step	Action in cycle of reference year T	Time
S1	Collecting data on legal units (T)	May T+1
S2	Validation of legal units (T): EGR IS	October T+1
S3	Collecting data on legal units (T), relationships (T) and enterprises (T)	November T+1
S4	Initial frame (T)	December T+1
S5	Structure validation (repair process) (T)	January T+2
S6	Preliminary frame (T)	February T+2
S7	Collecting data on groups (T)	February T+2
S8	Final frame (T)	March T+2

This paper describes the interplay between the Dutch SBR and the EGR, the use of the different EGR applications by Statistics Netherlands and address some quality issues of EGR information pointed out by Dutch FATS Statisticians.

2. Interplay between the Dutch SBR and the EGR

Since 2014, information about approximately 20.000 cross border relations were provided by the Dutch Chambers of Commerce (CoC). The identification process and the maintenance using the EGR-IS, meant a big increase of the workload of SBR staff. Besides, the statistical processes to exchange data with the EGR, depended strongly on the availability and the efforts of some specific staff members of the Dutch business register department. This might have affected not only the quality of the information provided by SN in the historical EGR cycles, but might also be a risk factor for future deliveries.

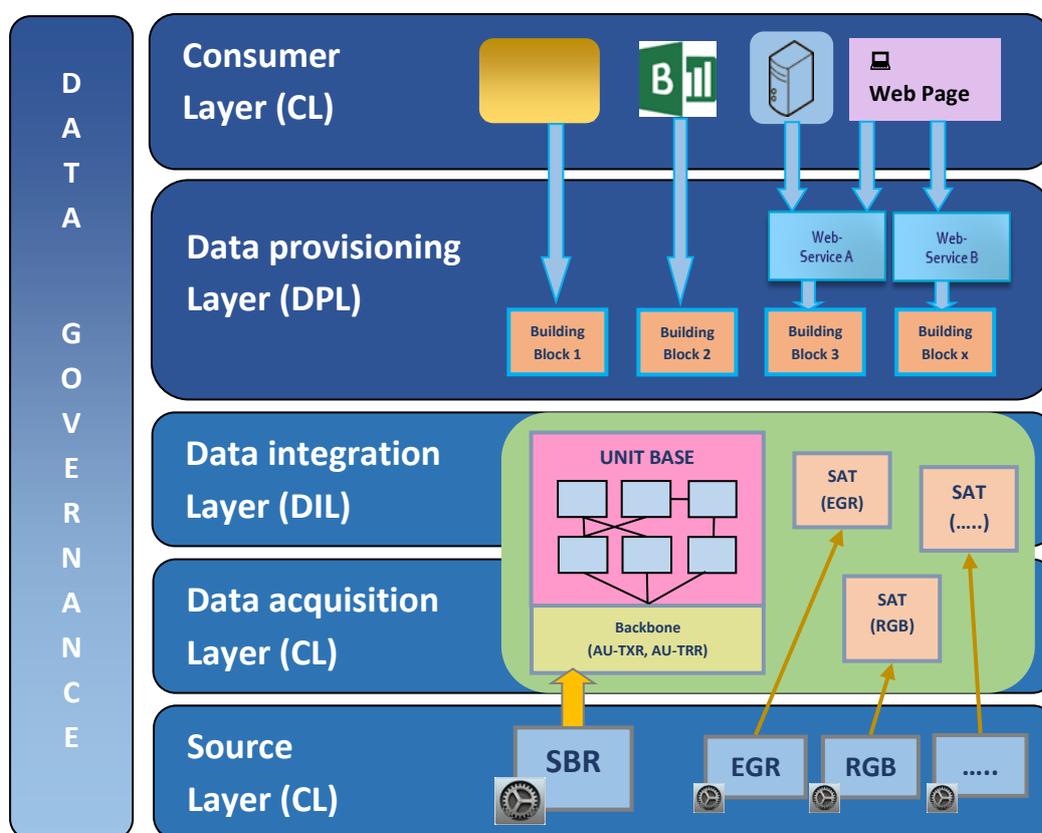
In order to integrate EGR methodology into the daily routines and stabilize the workload for SBR Staff and methodologists, SN developed in 2017 a special Cross-Border Register (CBR). The input of the CBR depends on the cross-border information gathered by profilers and the foreign ownership relations provided by the Dutch Chambers of Commerce in the Trade Register (TR). A central role in the maintenance of the SBR system is given by the data architecture that SN has adopted to reduce complexity of its SBR system. It explains how cross-order information is stored, updated and

distributed among its users. But before we will describe the technical infrastructure of the Unit Environment used by SN to support the different roles of the SBR.⁵

2.1 Unit Environment

Since 2017, SN introduced a new data architecture. This architecture describes the technical infrastructure to organize the different functionalities of the SBR in order to satisfy the requirements of the business regarding the ability to extract or consult microdata from the SBR at any moment in time. Therefore SN developed the Unit Environment. This environment is separated from the Live SBR and can be consulted by the users 24/7 by using different methods. This method depends on the aim of the user.

Figure 2: Data architecture of the Unit Environment



The concept of this data architecture is broader than the use of it within statistics Netherlands. The Unit Environment is used to reduce the complexity of the information stored in the SBR system and the linked micro data stored in other systems. The unit environment consists of several layers and it plays a crucial role to provide the input-information to the CBR.

Source layer (SL)	The SL comprises different sources and systems to alter and maintain information stored as different types of applications or document types in a Live Environment.
-------------------	---

⁵ Guidelines on Statistical Business Registers, chapter 2: <http://www.unece.org/index.php?id=40574>

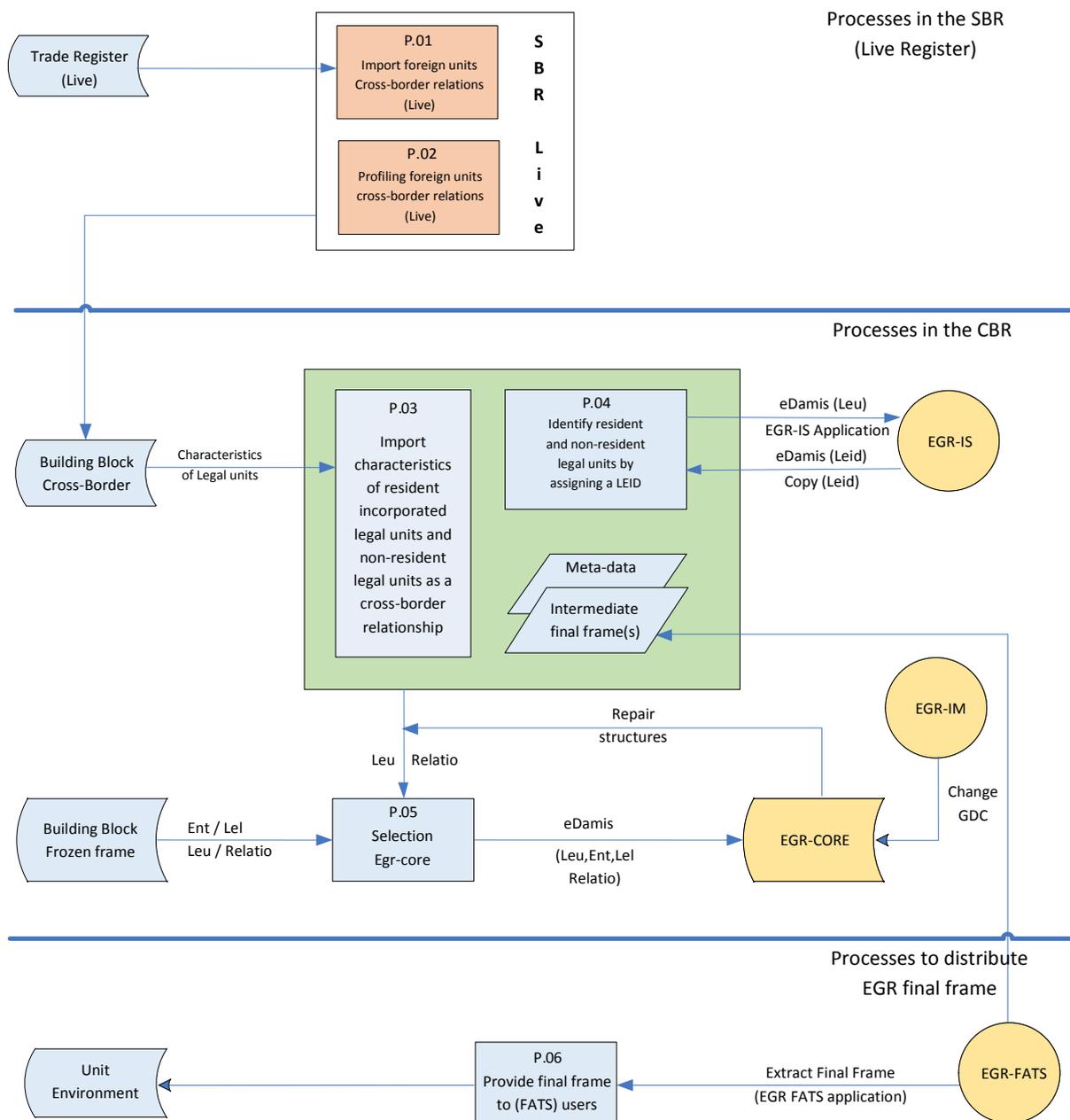
<p>Data acquisition layer (DAL)</p>	<p>The DAL allows data to be physically integrated into the system. The layer receives the data from the sources in the source layer. In the DAL the different types of characteristics are organized separately from each other, and even at different times.</p> <p>Each month a register snapshot is extracted from the SBR Live Register. This SBR Snapshot consists of</p> <ul style="list-style-type: none"> - the Unit Base (=frozen frame), which comprises the statistical units - the AU-backbone with the administrative units - the links between all types of units <p>Data from the other sources are extracted and stored as a satellite with its own time cycle. This means that the characteristics are stored as a frozen set of variables for a type of unit which occurs in the snapshot. E.g. the postal address, the email address or the phone number used to conduct a survey are stored in the SBR but are not stored in the Unit Base, because the users (data collectors) of this kind of information differ from the users (statisticians) which produce statistical indicators. At SN, each month the frozen frame is updated, but the contact information is updated each week, because of its specific use.</p>
<p>Data Provisioning Layer (DPL)</p>	<p>In the DPL, the information stored in the register snapshot is combined with the information stored in one or more satellites. The information is stored in building blocks that are accessible by the consumer layer. Each building block consists of a web service and primary keys to identify the unit type.</p>
<p>Consumer Layer (CL)</p>	<p>The CL consists of tooling to access the information stored in the building blocks in the DPL.</p> <p>With data preparation tooling (DPT) it is possible to combine the building blocks in the DPL with information of different types. It makes data accessible and combinable of different types, e.g. an Excel file, free text file, spss format or an xml-file with your own information.</p> <p>If you want to link a system of your own to the building blocks, then you can use a web service. The idea is to send a primary key to the web service and the characteristics stored in a building block are returned and are accessible to use in your own system. At SN, this kind of connection is used by the data collection department to use the information of the SBR in their own systems.</p> <p>SN has built a web interface with the web services to have the information in the building blocks consulted by multiple users.</p>

2.2 Cross Border Register

The CBR is a register and a system with functionalities to facilitate the identification process and the production of the required EGR datasets. The datasets will be checked with the EDIT-tool and sent with eDAMIS to the EGR. In the CBR the EGR-methodology is implemented in order to coordinate and ease the workload of the SBR staff. The system is developed by the IT department of SN. The CBR is

an extension to the SBR system⁶. The register department of SN is responsible for the maintenance of the SBR as well as for the CBR. After the CBR is taken into production, BR Staff will generate, send EGR files in the agreed format with eDAMIS or load EGR feedback files automatically into the CBR. It is not clear yet how much hours BR-staff will need for the maintenance of the CBR, scrutinizing populations or how much the work is involved in the manual identification process.

Figure 3: Process model Cross Border Register (CBR)



⁶ A description of the regular maintenance procures in the Dutch SBR is described in the [paper presented at the Wiesbaden Group in Vienna 2014](#).

Cross-border ownership relationships

If a Dutch legal unit is owned by a foreign incorporated entity, then the executives of the Dutch legal unit are obliged (according the Dutch Law) to have a 100% ownership relation registered in the Trade Register (TR) between the foreign incorporated entity and the Dutch Legal Unit. For each foreign incorporated entity, characteristics about its name, its location and the way it is registered in its home country are stored in the Dutch Trade Register of the Chambers of Commerce. This also includes information about the country and name of its national register and the national ID in it.

The micro data about the foreign entity and the ownership relation are stored as characteristics of the resident legal unit in the SBR Live Register⁷. Hence this information is accessible to be adapted by a profiler. In the SBR Live register a profiler is also allowed to enter a new cross-border relationship for any resident legal unit. The allocation of the LEID as characteristic of the foreign legal unit does not take part in the SBR Live Register, but in the CBR.

Importing cross border relations between resident and non-resident legal units in the CBR

The cross-border relation and the characteristics of each non-resident and resident incorporated legal unit, which are part of the cross-border relationship, are stored in the CBR. Each month the new set of legal units is compared with the last available set in the CBR. This means that each month the most recent set of cross-border relationships in the CBR is refreshed with entries, exits or updated characteristics. According the prescribed data architecture, the microdata is extracted from a special cross-border-satellite and a cross-border-building block, which include the cross-border relationships from the TR and those adapted or detected by the profilers. SN is planning to extend the functionalities of the CBR by compiling a 'virtual LEID' and produce a dataset with 'virtual LEIDs' which can be checked on validity by using the 're-identification service' of the EGR-IS (see "*Identifying non-resident legal units, 2.*").

Identifying resident legal units in the CBR

Identifying resident legal units with the EGR-IS requires a correct modification of the characteristics, also described in the guidelines "EGR IS resident legal units⁸". Important in this respect are the use of the correct set of characters and the application of the EDIT-tool⁹ to check your data before sending the file to the EGR-IS by using eDAMIS.

Selection of resident incorporated legal unit, which might be a part of a cross-border relationship:

1.	Legal unit with upward relationship with a non-resident legal entity
2.	Legal unit with downward relationship with a non-resident legal entity (subsidiaries)
3.	Legal unit with an institutional sectorcode in - S11003 Foreign controlled non-financial corporations

⁷ Guidelines on Statistical Business Registers: https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf

⁸ <https://webgate.ec.europa.eu/fpfis/wikis/display/EGR/Guidelines+1+-+EGR+IS+resident+legal+units>

⁹ <https://webgate.ec.europa.eu/fpfis/wikis/display/EGR/EDIT+validation>

	<ul style="list-style-type: none"> - S12xx3 Foreign controlled financial corporations - S99999 Entities with activities only in foreign countries
4.	<p>Legal unit with a 'foreign' legal form:</p> <ul style="list-style-type: none"> - 058 European public limited - 059 European cooperative public limited - 091 Formal foreign legal form inside of the European Union - 093 European Partnership - 094 Formal foreign legal form outside of the European Union - 095 Non formal foreign legal form outside of the European Union - 096 Foreign legal form - 099 European Group for Territorial Cooperation
5.	<p>Legal unit with a 'special' legal form:</p> <ul style="list-style-type: none"> - 012 General partnership (GP) - 025 Commandite partnership - 043 Limited Liability (LL) - 057 Corporation (Public Limited)
6.	<p>Legal units which belong to the 'TopX'</p> <p>The TopX is the set of most important EGs for Dutch Statistics ¹⁰. There is a high probability that the legal units that belong to these EGs have relationships abroad.</p>
7.	<p>Rest of the legal units which belong to the domestic EG (DEG) together with one of the legal units selected in 1-5</p>

The sets of legal units offered to the EGR-IS by eDAMIS are based on the reference periods of December T. The legal units from the previous reference year of December T-1 that have ceased their activities are also included in the selection, but the closing date is filled with the aim to prevent those from being selected in the new final frame. The selection script is a special Use Case developed in SQL programming language as a Stored Procedure. Adjustments are applied to the register data to fit the EGR rules, like the date format, and special characters are changed to authorized characters (e.g. Ã to A, Æ to AE, Ç to C).

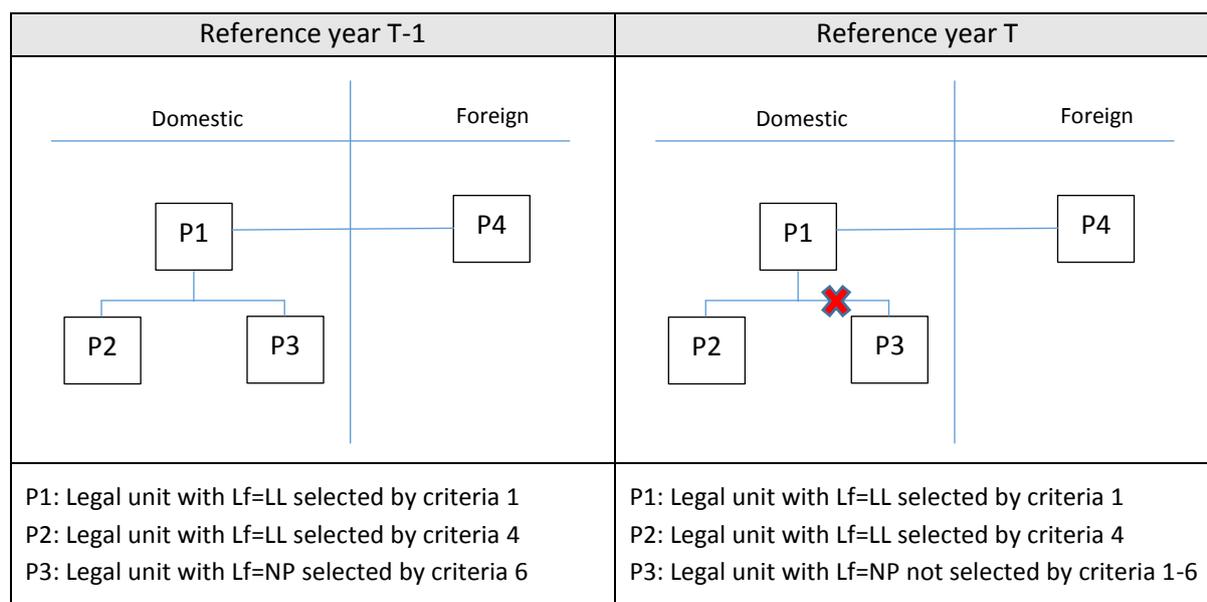
Problem in the selection-procedure for the EGR-IS

Eurostat expects that SN provides information to the EGR-IS about all (potential) Dutch incorporated legal units which are or might be involved in a valid cross-border relationship in the reference year, complemented with information about liquidated units which were valid in the previous reference year. But neither is the case, when the selection procedure on behalf of the EGR-IS excludes legal units with specific legal forms, which are supposed to be not incorporated. Natural persons, private equity firms and most governmental bodies (defined by legal form) are excluded from this selection, no

¹⁰ A description of the regular maintenance procures in the Dutch SBR is described in the [paper presented at the Wiesbaden Group in Vienna 2014](#).

matter if they are involved in a cross border relationship or not. However there is one exception to this rule, e.g. if a natural person (NP) has entered the selection according to the criteria described in point 6. But a problem may rise when the ownership relation with the still existing natural person is not valid anymore for the next reference year (see example in the visualization). In that case the natural person will not be a part of the ‘incorporated’ set, e.g. because of the legal form restriction according selection criteria 4 (see drawing). Hence it should be removed from the EGR-IS dataset, although the natural person is still an active legal unit. But if it is removed from the EGR-IS, a profiler is not able anymore to identify it by using the Interactive Profiling Tool (IPT).

Visualization of a problem in the selection-procedure for the EGR-IS, caused by broken relationships



Identifying non-resident legal units

Since there are about 20.000 cross-border relationships with non-resident legal unit to be identified (or validated) with the EGR-IS each reference year, it is not practical to do this manually (especially the first time). The LEID of successfully identified (foreign) legal units will be stored in the CBR. At this moment, the identification process is a manual process and not yet implemented automatically as a feature from the CBR.

The following methods are applied to find a correct LEID of a non-resident legal unit:

1. Previous LEIDs	For the EGR cycle of the last reference year, many non-resident legal units were (manually) identified by using the EGR-IS. In this step the validity of those cross-border relationship is checked for the actual reference year by comparing the names of the non-resident legal units involved. If a resident legal unit has a cross-border relation in reference year T-1 and an actual relationship in reference year T as registered in the TR, then the names of the non-resident legal units are compared on the corresponding characteristics in the name by using a ‘trigram’ method. If the overlap in
-------------------	---

	the name is sufficient, then the relationship is also qualified for the actual reference year (and the LEID of T-1 is also valid for T).
2. Virtual LEIDs	The identification process in the EGR-IS is supporting the validation of a list of known LEIDs. This functionality requires a list of LEIDs as input. The output of this batch process provides information about the validity of the LEID. For this process, SN derives a virtual LEID with the information provided by the TR about the national register and the national ID of the non-resident legal unit. For each participating country in the EGR, the formats of the National IDs in the EGR-IS ¹¹ are used to construct a virtual LEID. This list of virtual LEIDs are offered to the EGR-IS batch process to check the validity.
3. Name and address	This application describes the standard use of the EGR-IS batch process. A list of names and addresses of the non-resident legal unit is send to the EGR-IS. This application uses its link-algorithm to search the perfect fit for the non-resident legal unit in the EGR-IS database.
4. Manual checking	Manual checking by using the EGR-IS application is an intensive process. To optimize this process a list of non-resident legal units in a cross-border relationship is which is not yet detected by applying one of the methods above. The cross-border relationships are ordered by importance of the (domestic) Enterprise Group of the resident legal unit (e.g. employment). The non-resident legal units in the list are top-down manually identified. If the delivery to the EGR-CORE is due, then this identification process will stop.

Delivery of EGR-CORE microdata to the EGR

The delivery to the EGR-CORE comprises a set of LEUs, ENTs, LELs and RELs.

- (1) The sets are based on the information about those resident legal units (LEUs) who are involved in a cross-border relationship. For the record, in this selection the non-resident owners do not have to be completely identified yet by a LEID.
- (2) The set of national enterprises (ENTs) consists of those enterprises with a link to at least one LEU in the set of LEUs selected in (1).

¹¹<https://webgate.ec.europa.eu/fpfis/wikis/download/attachments/130162211/NATIONAL%20ID%20formats%20in%20EGR%20IS%20V06%20June%202018.doc?api=v2>

- (3) The set of LELs comprise all links between an ENT in the set of the ENTs in (2) and all resident legal units linked to it. This means that this set contains also legal units which are not included in set (1).
- (4) The last set of the RELs, comprises all downward resident relationships and all upward cross-border relationships of the legal units in (1) and which are completely identified. This means that from each cross-border relationship the LEID of a resident legal unit in (1) is known as well as the linked LEID of the non-resident legal unit.

Again, it is advised to use the application of the EDIT-tool to check your data before sending the files to the EGR by using eDAMIS.

Notes:

1. If in the regular delivery of the EGR-CORE, necessary information about some cross-border relationships is missing which involve a resident legal unit, the EGR may request for additional set with the required EGR-CORE information. In this case the CBR has the functionality to compose a new delivery of the EGR-CORE based on an additional set of resident legal units. This functionality is adopted by SN to repair broken structures (REL) and complete the information in the EGR about Dutch units involved in Global Enterprise Groups (GEG). Hence for this purpose we do not use the EGR-IM (see paragraph 3 on how the EGR-IM is used by SN).

Besides this gives SN also the opportunity to complement the Dutch OFATS population by providing information from new sources to the EGR, like information about small & medium ENTs with Extra-EU affiliates companies' collected by the tax office (downward links). In the Dutch SBR, domestic EGs with the largest impact on statistics are (manually) maintained and monitored by profilers¹². This work also includes the validation of cross border relationships with each legal unit belonging to these EGs.

2. Relationships with Dutch legal units which are delivered by another NSI in the previous final frame are not used by SN as input in the CBR. This is not preferred, because then CBR would become the authentic source for these kind of relationships and we would have no source for following up on these relationships.

Use of the EGR-IM

The EGR IM is an interactive and a useful tool. It displays the structure of enterprise groups, provides insight into the levels of the units and allows the user to perform the repair action. The most common usage of the EGR IM is to add some new legal units to existing enterprise groups and edit the relationship between these legal units and other parts of enterprise groups, whether this relationship is a parent or a child. It is also useful to remove some non-existing relationships.

¹² A description of the regular maintenance procedures in the Dutch SBR is described in the [paper presented at the Wiesbaden Group in Vienna 2014](#).

Another advantage of the EGR IM tool is that authorized BR staff may manually assign another legal unit as the GDC or update the attributes of a GEG, such as the country code of the GDC, the NACE code, the number of persons employed, turnover, etc.

SN will explore the use of the EGR-IM in more detail in a pilot-project granted by Eurostat: Demonstration of interactive and collaborative EGR production using EGR IM and resolution of governance issues.

SN expects to use the EGR-IM to

- Visualize the relationship structures of the groups, with focus on cross-border relationships
- Compare the group structures with the group structures in our national BR
- Validate relationships and identify broken structures
- Update the information about the GDC of a GEG

Handling EGR frames

When each NSI provides its cross-border information by the EGR-CORE delivery, has completed their repair actions and has validated their GEGs with a resident relationship, then the EGR can compose a final frame about the reference year (15 months later). In the meantime, the EGR produces an initial and a preliminary frame to be used by the national register staff to raise the quality of the characteristics stored in the EGR.

The preliminary frame is used by several countries to extract those GEGs, with a resident Global Decision Centre (GDC¹³) in order to support the data collection of the Outward FATS statistics. The final frame is used to report on IFATS and OFATS statistics according the EU-regulations.

Each frame consists of 3 sets of data about legal units (LEU), Enterprises (ENT) and Global enterprise Groups (GEG), see information model in the [Annex](#). The sets of the EGR final frame are not recorded in the CBR, but is distributed to the users as 3 special building blocks.

3. Improvements for Dutch FATS statistics

FATS Statistics describe the influence on the economy of one country on another country by allocating foreign affiliates to that country where the ultimate decisions on production are made, the so called country of the Ultimate Controlling Institute (UCI). In the EGR the UCI is called the Global Decision Centre (GDC). After the composition of EGR final frame, any national FATS-statistician can extract that part of the frame with the EGR-FATS interface, in which a legal unit of his country is participating. It would have been a less complicated job to produce harmonized FATS statistics, if the GDC assigned by the EGR was of 'good quality'. However the sources and the EGR methodology to allocate the country of the GDC are not sufficient enough yet to determine the correct country of the GDC for its purpose. According to the recommendations on the GDC¹⁴, the GDC in the EGR is either identified in the EGR by a correct or an incorrect LEID or it is not identified in the EGR due to missing upward relationships.

¹³ The GDC in the EGR equals the Ultimate Controlling Unit (UCI) of FATS.

¹⁴ <https://webgate.ec.europa.eu/fpfis/wikis/display/EGR/Recommendations+on+the+GDC>

A wrong country code for the GDC in the EGR implies that FATS Statisticians will improve the quality by themselves and that the EGR will just be an important source to point out a possible country code for the GDC out of other sources. The next paragraph describes the methodology applied by Dutch FATS statisticians to improve the country of the GDC for FATS.

3.1 GDC located outside the EU

The country of the GDC for Dutch FATS statistics is determined for national ENTs stored in the Dutch SBR. The GDC is determined for the 31st of December of the reference year as subscribed by the FATS manual. The delineation process is split into an automatic and a manual investigation process to determine the GDC.

Automatic process

The automatic process starts by gathering information about the legal structure of the national ENTs that were active during (part of) the reference year T in the specific NACE-groups of FATS. Next, the country code of the GDC of the corresponding legal units of the previous reference year T-1 is added as a characteristic to the set. At SN several internal and external sources are available which provide information about a proposed country of the GDC. The country code of the GDC found in each source is also linked to the national ENTs.

1.	Survey Finances Large Enterprise Groups (SFEG) (Unit type = DEG)	This survey covers profit and loss information as well as stock positions at the beginning and end of the year and financial transactions (purchases and sales). Important for this purpose is the specification to domestic and abroad.
2.	EuroGroups Register (EGR) (Unit type = national ENT)	The EuroGroups Register (EGR) is the statistical business register of multinational groups in the EU. The EGR contains structural economic information on multinational groups with an interest in Europe and on their enterprises.
3.	Netherlands Foreign Investment Agency (NFIA) (Unit type = Legal Unit)	The NFIA is an operational unit of the Dutch Ministry of Economic Affairs. It supports foreign affiliates which want to establish or extend activities on Dutch Territory. The NFIA provides information to these companies about taxes, legislative procedures and business locations in the Netherlands. Besides the NFIA is responsible for the international promotion of the Netherlands as being a country for entrepreneurs and international businesses.
4.	TopX ¹⁵ - relationships (TOPX) (Unit type = ENT)	In this source, information is stored about the GDC or the foreign affiliates of Large and Complex EGs which belong

¹⁵ A description of the regular maintenance procedures in the Dutch SBR is described in the [paper presented at the Wiesbaden Group in Vienna 2014](#).

		to the TopX. The information is collected 'manually' by profilers and Concern Coordinators.
5.	Trade Register of the Chamber of Commerce (TR) (Unit type = Legal unit)	The TR is the main source of legal units of the Dutch SBR. The TR also stores information about the upward foreign relationships of Dutch legal units.
6.	Survey Investments in Tangible Goods (INIVA) (Unit Type = ENT)	The survey describes the realized investments ¹⁶ in tangible goods and the net realizable value of the sale of tangible fixed assets

The prediction of the country of the GDC by a source are prioritized by a specific weight which depends on the country: The Netherlands (NL), Luxemburg (LU), Malta (MT), Cyprus (CY), Switzerland (CH) or the United Kingdom (UK) are popular countries chosen by GEGs with a controlling unit located outside of Europe to establish an Offshore Financial Centres (OFC). According the FATS-Manual¹⁷, "units without a substantial physical presence and without significant economic activities cannot engage in decision-taking or effectively manage subsidiaries. This applies to 'empty shell' units located in tax havens and in offshore financial centres, to special-purpose entities and to many non-profit units (e.g. foundations, trusts, 'Stiftungen', 'Anstalten'). Thus, they should not be appointed as UCIs".

If a source reports a GDC located in one the countries mentioned above, then the GDC is prioritized as last, which means that it can only be pointed out as GDC, if there is no other source that provided information about a GDC located in another country. Each source may provide conflicting information. This is caused by the interpretation of the respondent, who answered the question about the location of the decision centre or who represents the 'mother' of the GEG. In many cases the respondent is only aware about the direct foreign owner, but does not know the other upward links.

The information from the sources above is linked to the ENTs by different unit types. If we exclude the different characteristics about names, addresses, activities, employment, etc. then the basic record of an arbitrary DEG might look like this:

DEG-ID	ENT-ID	NSA-ID	T-1	SFEG	EGR	NFIA	TOPX	TR	INIVA
12345678	11111111	00000001	US	NL	LU	-	UK	-	-
12345678	11111111	00000002	US	NL	LU	US	UK	-	-
12345678	11111111	00000003	US	NL	LU	-	UK	-	-
12345678	22222222	00000004	US	NL	-	-	-	-	CH
12345678	22222222	00000005	US	NL	-	-	-	LU	CH

The automatic investigation of the sources results in an initial resident or a non-resident GDC according the priority order as presented. This means that if for example the SFEG reports US and the EGR reports the Netherlands (NL) as the GDC, then the first guess to be the GDC will be US and not

¹⁶ Investments in intangible and financial assets are excluded.

¹⁷ <http://ec.europa.eu/eurostat/documents/3859598/5922981/KS-RA-12-016-EN.PDF/c93cdf48-5efa-459f-b218-731a9a5476e9?version=1.0>

NL, because the SFEG is prioritized higher as a source than the EGR is. If no information about a foreign GDC (i.e. <> NL) is reported, then the GDC is assumed to be Dutch (=NL) in the first stage of the delineation process.

Manual process

The manual process is all about providing the best educated guess about the country of the GDC. It often depends on the picture of the structure of the GEG you may obtain by googling on the names of the subsidiaries and look for the direct upward link until a proof for the correct country of the GDC is found.

The process is triggered by the information provided by the sources and a ranking by importance based on the persons employed. There are several triggers to manually investigate the country of the GDC. Each year about 3000 DEGs are checked manually by Dutch FATS statisticians. In the table below, the most important triggers to manually check the GDC of a DEG are listed.

The GDC has changed compared to the previous year	This might be a trigger for a big reorganization of the global firm, which have a big impact on FATS and other statistics.
Several sources report different GDCs (not being an offshore financial centre) for the same DEG.	This might have caused some confusion to point out the right location of the GDC. Investigations might give the correct information.
The sources only report an OFC as the GDC	This might be very appropriate for the Netherlands, since there are many Special Purpose Entities registered in the Dutch SBR. In most cases the location of an OFC might not be the location of the GDC.
the GDC is reported as Belgium (BE), Luxemburg (LU), Switzerland (CH), Malta (MT) or Cyprus (CY) and has not been manually checked in the past.	The countries listed host many financial intermediates. Extra research should point out if the country of the GDC is pointed out correctly by the automatic process.
the GDC is not NL and it is not a corporation	If the GDC is not a corporation, then we might have to check if the GDC assigned, is the same as the Ultimate Beneficial Owner or an intermediate entity.

The manual checking process accounts for the knowledge as listed below. The suggestions are based on the experiences of Dutch FATS statistics. In some cases the checking procedures are only applicable for Dutch units.

- Research on the structure of the enterprise group gives information about the legal units and their legal names. The legal name often comprises foreign phrases of the original country where the GDC is located (Like Siemens Luxembourg S.A.R.L.).
- Googling the name of an arbitrary subsidiary often leads to <https://opencorporates.com/>. This website provides information about the organization and the structure of many GEGs, including the link to the company's website. It often occurs that a GDC outside of Europe will open a corporate as a Luxembourgian Limited, recognized with the wording 'SARL' in its legal name.
- The corporate structures of American GEGs are often reported on <https://www.sec.gov/> or <https://www.bloomberg.com/europe>. American GEGs often establish a legal entity in Ireland (IE) to support their activities in Europe.
- Israeli GEGs often established a legal entity in Malta (MT) to support their activities in Europe, while Russian firms often own a legal entity in Cyprus for this purpose.
- According to Dutch law, a board member of an incorporated enterprise might be held responsible for his or her policy when the business cannot fulfil its obligations regarding the payments of taxes and contributions as a consequence of mismanagement¹⁸. Therefore some leading board members will establish an official legal holding entity in England to by-pass this risk. These entities can often be recognized by the addition 'Limited (LTD)' in their legal name. In these cases it is not likely that the country of the GDC is the UK, but NL. Googling the name of this particular board member often results in a Dutch home address, which more or less proves that the board member operates from NL.
- If a source assigns Belgium (BE) or Germany (DE) as the country where the GDC is located, then more information might be useful because of these are neighbouring countries of NL.
- Sometimes the SFEG or another source reports a Dutch cooperation to be the GDC, but one of the 'members' of the cooperation is from the US. In this case it is more likely that the GDC is located in the US.
- On the company's website, the company's history often provides information about the country where the firm was founded. Due to fiscal or legal reasons the firm might have decided to organize and register the global structure in different countries.
- The location of the headquarters may betray the main location of the members of the board and where the most decisions are taken.
- The monetary unit of the annual report provides information of the most important currency of the 'home country'.

¹⁸ See Dutch law about "Bestuurdersaansprakelijkheid":

https://www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/belastingdienst/zakelijk/aangifte_betalen_en_toezicht/aansprakelijkheid/bestuurdersaansprakelijkheid/

- The annual general meeting is an important instrument for the board of the firm to communicate with the responsible personnel of its subsidiaries. The location of this meeting gives information about the location of the GDC.
- The contact information of investors relations, may consist of addresses and telephone numbers which point in the direction of the location of the GDC.
- It is useful to consult the partnering countries of a GEG. The colleagues of the partnering country might have more information about location of the GDC.

Some Dutch legal units fulfil a more administrative function in specific industries, like in real estate, but are owned by a foreign GDC. These cases should be investigated manually because of their importance to describe the influence of the foreign country in the industry, but there is no employment. Hence those cases might never be selected for manual investigation, because their importance is quantified by other criteria, e.g. their balance sheet total instead of the number of people working.

Regularly, manual investigation is not triggered if the country of the GDC assigned by the EGR for reference year T-1 is the same as the country code assigned by the EGR for reference year T, because the impact for FATS is minimal. By coincidence, cases are detected where the country code is correct, but the LEID of the GDC assigned by the EGR does not point to the GDC found when the structure is investigated manually.

At SN the checked list about the GDCs assigned by the FATS Statisticians is final, but also used by the statisticians of the International Trade in Goods Statistics (IGTS). Sometimes anomalies are detected when transactions are analysed in the ITGS. These anomalies may lead to adapt the 'final' list.

A Dutch FATS Statistician with enough experience, may recover the GDC by investigating 15-30 DEGs manually per day by applying this methodology. This depends on the complexity of the GEG.

3.2 Developments

During the last EGR workshop in September 2018, developments on the EGR system were presented and discussed by the EGR users. The improvements on the EGR were collected from the EGR team and from users in the NSIs. They comprise improvements in the EGR applications which help improving the quality of the EGR information and provide new functionalities to support the users to verify it. In the future it might seem useful to develop a kind of messaging system to share changes in the GDC information applied by the users of the NSIs involved, for validation.

EGR-IM	<ul style="list-style-type: none"> - Include data of input and throughput quality indicators in EGR IM user interface - Improve the GEG tree visualization
EGR-IS	<ul style="list-style-type: none"> - Improvement of search algorithm in EGR-IS - Add meta-information on the identification results in the EGR-IS (e.g. GDC information from previous year T-1) - Avoid having duplicates in EGR IS which lead to inconsistency in the group structures

EGR-CORE	<ul style="list-style-type: none"> - Improvement of Group Continuity Algorithm - Provide a report per NSI on non-consolidated relationships - Special Purpose Entities (SPEs) not to be assigned as GGH/GDC. They may wrongly change group nationality
EGR-FATS	<ul style="list-style-type: none"> - Improve EGR-FATS Interface to include visualization tab for the tree structure of the legal units of the groups - Include data of output quality indicators in EGR FATS user interface

However the aim of the EGR is to provide frame populations that can be used to produce coordinated FATS statistics within the EU. For the OFATS users, this means that the identification and the characteristics of the GDC as a unit should be stored correctly in the EGR, because it should be possible that SN can conduct an extra survey to collect extra information. For the production of IFATS indicators it is important that the country of the GDC for each Dutch enterprise in the population - which describe the Structural Business Statistics (SBS) - is coordinated and of good quality. This last requirement can be achieved when the country of the GDC assigned by Dutch FATS statisticians, is visible in the preliminary and final frames, extracted from the EGR (e.g. by using the EGR-IM to correct the country of the GDC).

SN would like to improve the quality of the GDC for Dutch legal units under foreign control, especially if the GDC is located outside the EU. The suggested methodology to assign the country of the GDC can be seen as a first attempt to support IFATS users. It is based on the use of google and the information published on specific websites to investigate GEG-structures. During the investigation process legal names, addresses, and national identifications of subsidiaries and the GDC are visible on the internet. It would be thinkable to apply web-scraping and text mining methods in the near future to create a data repository and to develop an ontology in order to support the 'manual part' of the identification process.

For SN, it is an important step to centralize the EGR methodology in the CBR in order to coordinate the Dutch deliveries to the EGR and to optimize the use of the EGR applications within SN. In the future, SN would like to embed the suggested FATS methodology in the CBR to improve the quality of the country of the GDC assigned by the EGR and support the production of Dutch FATS statistics.

Annex: Information model EGR-frame

NEW GROUP EXPORT FILE - from April 2017	
Field name	Field description
FRAME_RYEAR	Frame reference year
FRAME_VERSION	Frame population version
GEG_EGR_ID	Group EGR ID
GEG_NAME	Group name
GEG_GGH_LEID	Global group head (GGH) LEID
GEG_GGH_EGR_ID	GGH EGR ID
GEG_GGH_NAME	GGH name
GEG_GGH_COUNTRY_CODE	GGH country code
GEG_UCI_LEID	Global decision centre (GDC) LEID
GEG_UCI_EGR_ID	GDC EGR ID
GEG_UCI_NAME	GDC name
GEG_UCI_ADDRESS	GDC address
GEG_UCI_CITY	GDC city
GEG_UCI_POSTAL_CODE	GDC postal code
GEG_UCI_COUNTRY_CODE	GDC country code
GEG_UCI_COUNTRY_CODE_T_1	GDC country code year T-1
GEG_NACE_CODE_DIV	Group NACE at 2-digit level
GEG_NACE_CODE_DIV_T_1	Group NACE year T-1 at 2-digit level
GEG_PERS_EMPL	Group employment (persons employed)
GEG_TURNOV	Group turnover in million
GEG_TURNOV_CUR_CODE	Group turnover currency code
GEG_T_ASSET	Group total assets in million
GEG_T_ASSET_CUR_CODE	Group total assets currency code
GEG_N_LEU	Number of legal units in group
GEG_N_ENT	Number of enterprises in group
GEG_N_COUNTRY	Number of countries with legal units
GEG_SIZE	Group size
GEG_COMPL	Group complexity
GEG_EU_INFL	Group European influence
GEG_EU_IND	Group EU indicator
GEG_DESCR	Group written description

NEW ENTERPRISE EXPORT FILE - from April 2017	
Field name	Field description
FRAME_RYEAR	Frame reference year
FRAME_VERSION	Frame population version
ENT_EGR_ID	Enterprise EGR ID
ENT_NSA_ID	Enterprise NSA ID
ENT_NAME	Enterprise name

ENT_ADDRESS	Enterprise address
ENT_CITY	Enterprise city
ENT_POSTAL_CODE	Enterprise postal code
ENT_COUNTRY_CODE	Enterprise country code
ENT_NACE_CODE	Enterprise NACE
ENT_NACE_CODE_T_1	Enterprise NACE year T-1
ENT_PERS_EMPL	Enterprise employment (persons employed)
ENT_TURNOV	Enterprise turnover
ENT_TURNOV_CUR_CODE	Enterprise turnover currency code
ENT_INST_CODE	Enterprise institutional sector
ENT_EU_IND	Enterprise EU indicator
ENT_N_LEU	Number of legal units in ENT
GEG_EGR_ID	Group EGR ID of the group of the enterprise
GEG_NAME	Group name
GEG_GGH_COUNTRY_CODE	GGH country code
GEG_UCI_LEID	Global decision centre (GDC) LEID
GEG_UCI_EGR_ID	GDC EGR ID
GEG_UCI_NAME	GDC name
GEG_UCI_COUNTRY_CODE	GDC country code
GEG_UCI_COUNTRY_CODE_T_1	GDC country code year T-1

NEW LEU EXPORT FILE - from April 2017	
Field name	Field description
FRAME_RYEAR	Frame reference year
FRAME_VERSION	Frame population version
LEU_EGR_ID	Legal unit EGR ID
LEU_LEID	Legal unit LEID ID
LEU_NAT_ID	Legal unit national ID
LEU_NAT_ID_NIS_CODE	Legal unit national ID NIS code
LEU_NSA_ID	Legal unit statistical ID
LEU_VAT_ID	Legal unit VAT ID
LEU_DUNS_ID	Legal unit DUNS ID
LEU_BVD_ID	Legal unit BVD ID
LEU_NAME	Legal unit name
LEU_ADDRESS	Legal unit address
LEU_CITY	Legal unit city
LEU_POSTAL_CODE	Legal unit postal code
LEU_STATE_CODE	Legal unit state code
LEU_COUNTRY_CODE	Legal unit country code
LEU_COUNTRY_NAME	Legal unit country name
LEU_EMAIL	Legal unit email
LEU_TYPE	Legal unit type
LEU_STA_CODE	Legal unit status code

LEU_LFORM	Legal unit legal form
LEU_SPE_CODE	Legal unit SPE code
LEU_NACE_CODE	Legal unit NACE
LEU_PERS_EMPL	Legal unit employment (persons employed)
LEU_DATE_INC	Legal unit date of incorporation
ENT_EGR_ID	Enterprise EGR ID
ENT_NSA_ID	Enterprise NSA ID
REL_P_LEU_EGR_ID	EGR ID of the parent legal unit
REL_P_LEU_LEID	LEID of the parent legal unit
LEU_GLOBAL_LEVEL	Level of the legal unit in the enterprise group
GEG_EGR_ID	Group EGR ID of the group of the enterprise
GEG_NAME	Group name
GEG_GGH_COUNTRY_CODE	GGH country code
GEG_UCI_LEID	Global decision centre (GDC) LEID
GEG_UCI_EGR_ID	GDC EGR ID
GEG_UCI_NAME	GDC name
GEG_UCI_COUNTRY_CODE	GDC country code
GEG_UCI_COUNTRY_CODE_T_1	GDC country code year T-1

List of Abbreviations

Abbreviation	Description
BvD	Bureau van Dijk
CBR	Cross-Border Register
CoC	Chambers of Commerce
D&B	Dunn & Bradstreet
DEG	Domestic Enterprise Group
eDAMIS	electronic Dataflow Administration and Management Information System
EGR	EuroGroups Register
ENT	(National) Enterprise
ESBRs	European System of interoperable Business Registers
ESS	European Statistical System
FATS	Foreign Affiliated Trade Statistics
GDC	Global Decision Center
GEG	Global Enterprise Group
GGH	Global Group Head
IFATS	Inward FATS
IGTS	International Trade in Goods Statistics
INIVA	Survey Investments in Tangible Goods
IPT	Interactive Profiling Tool
IT	Information Technology
LEID	Legal Entity IDentifier
LEL	Link between Enterprise and Legal unit
LEU	Legal Unit
MNE	MultiNational Enterprise Group
NCB	National Central Bank
NFIA	Netherlands Foreign Investment Agency
NP	Natural Person
OFATS	Outward FATS
OFC	Offshore Financial Centres
REL	Relationship between Legal units (domestic and cross-border)
SBR	Statistical Business Register
SBS	Structural Business Statistics
SFEG	Survey Finances Large Enterprise Groups
SN	Statistics Netherlands
SPE	Special Purpose Entity
SQL	Structured Query Language
TOPX	List of (important) Dutch Enterprise Groups maintained by profilers
TR	Trade Register
UCI	Ultimate Controlling Institutional unit