



**Economic and Social  
Council**

Distr.  
GENERAL

ECE/CES/GE.42/2009/4  
28 July 2009

Original: ENGLISH

**ECONOMIC COMMISSION FOR EUROPE**

**STATISTICAL COMMISSION**

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Business Registers

Eleventh session

Luxembourg, 6-7 October 2009

Item 4 of the provisional agenda

**SOURCES FOR IDENTIFICATION OF ENTERPRISE GROUPS (BEFORE PROFILING)**

**Data Sources of the Eurogroups Register**

Note by Eurostat

**I. DATA OF THE EUROGROUPS REGISTER**

1. The EuroGroups Register (EGR) is the statistical register of the European Communities on multinational enterprise groups (MNEs). The aim of the register is to cover the units of the most influential MNEs in the European Union (EU) and the European Free Trade Association (EFTA) countries to create a harmonized frame for European statistics on multinationals.
2. The EGR operates in yearly cycles; the present production phase is the cycle 2009. In this cycle the data of multinational enterprise groups for reference year 2008 are processed in the register. The national statistical institutes (NSIs) of the EU and EFTA countries are involved in the EGR production process. The NSIs link the records of the commercial data sources to their national business registers; they check and correct the characteristics and the relationships of the units at record level. After the amendments of the Member States the central EGR system defines the enterprise group structures.
3. This document presents the commercial data sources of the EGR, the composition of the starting population, the characteristics and the completeness of the initial data sets. The results of the linking to national business registers of Member States will be presented as well.

GE.09-23507

## **II. DATA ACQUISITION FROM PRIVATE PROVIDERS**

4. In 2008 the Statistical Office of the European Communities (Eurostat) announced a tender for private data providers to supply data on multinational groups for statistical purposes to the EuroGroups Register. The tender was slated to two lots. One lot requested data on control relationships, the other one data on ownership relationships.

5. The tender specification required data of 5000 MNEs with the characteristics of all involved subsidiaries and the relations between them. For the dataset on control relationships the winner of the tender was company Dun and Bradstreet (D&B), for the dataset on ownership relationships company Bureau van Dijk (BvD). Both of these private data providers do have a worldwide commercial database on multinational groups.

## **III. SELECTION PROCEDURE FOR GROUPS**

6. By the definition of the 5000 starting population the purpose of Eurostat was that EGR should cover the most influential MNEs active in Europe, those which are the most relevant from the perspectives of Foreign Affiliates Trade in Services (FATS) statistics, Foreign direct investment (FDI) statistics and employment.

7. The selection methodology for the starting population of the EGR was defined by Eurostat. As a first step Eurostat asked the EU and EFTA Member States to provide lists on influential multinational enterprise groups which are resident in their countries and lists on MNEs which are the most relevant for their FATS and FDI statistics. Eurostat predefined the numbers of requested MNEs by countries. The intention was to cover a large number of groups from large economies, e.g. United Kingdom, Germany, France, but even very small countries e.g. Lichtenstein or Malta were asked to propose some groups for the initial EGR population.

8. From the proposals of the Member States Eurostat compiled a list with more the 3000 supposed group heads. The group heads were identified with names, addresses and with national identification numbers received from the Member States.

9. In October 2008 Eurostat provided 3072 supposed group heads for both private data providers to match them to their database. The matching was based on the national identification (ID) numbers and on the names of the group heads. Some countries, which had direct relations to the private data providers at national level, could provide Data Universal Numbering System (DUNS) numbers and BvD numbers as well. These numbers are the identification numbers of the private sources and made the matching simple. D&B could match 2410 groups to its database; BvD could match 2729 groups. These matching units composed the first half of the starting population.

10. In addition of the matched groups Eurostat asked the private sources to select additional groups for the EGR. First we asked a country specific selection, the number of requested groups by countries were defined by Eurostat. The private sources provided around 1500 MNEs following the needs of Eurostat from country to country. After that the private sources provided further important European and non-European groups, irrespectively to their country to top up

the 5000 starting population. These MNEs only were based on the biggest global employment for European groups and on the biggest employment in Europe for non-Europeans.

11. D&B totally provided the data of 5000 multinational enterprise groups and their constituent units, BvD provided the data of 5452 MNEs, of which 452 records were additional "bonus" to the original contract.

#### **IV. THE DELIVERED DATA FILES FROM PRIVATE SOURCES**

12. D&B provided the final delivery file to Eurostat in January 2009. The dataset had the time stamp of 31 December 2008; this means that the units, the relationships and all values of the variables reflected the status of the D&B database at the end of 2008. The final delivery included 3 files. The main data file included 330 296 records; these records are the legal units of the 5000 MNEs of the starting population.

13. D&B provided a separate file on the list of the 5000 MNEs as well. Alongside these files D&B also provided a file which included consolidated global numbers for the MNEs on employment, net worth, turnover and statement date for financials. The format of the data files met the specifications of Eurostat.

14. BvD provided the final delivery file at the same time with the same time stamp. The final delivery included data of 5452 MNEs. The data were provided in 3 files. The main data file, called "Entities" included 505 560 records; these records were the legal units of the starting population. BvD provided a separate file on the parent-subsidiary relationships of the legal units, called "Links" with 588 505 records. Alongside these files D&B also provided a file which included consolidated global numbers for the MNEs. The format of the data files met the specifications of Eurostat.

#### **V. COMPLETENESS OF THE DUN AND BRADSTREET DATA**

15. The main D&B data file included the variables of the legal units of the MNEs and the relations of the units to immediate parents and to the global group heads as well. For the 330 296 records the delivered file included 52 variables.

##### **A. Identification number**

16. Every record had a different DUNS number as unique identifier in the proper 9-digit format with preceding zeros if needed. The missing preceding zeros could make troubles as they did in the EGR test deliveries, it is useful to start all numeric identifiers with a number which is not 0. The internal ID numbers of EGR always start with 1.

**B. National identification numbers**

17. These numbers were very important for the linking process between the two private sources. 80 per cent of the records did have national ID number. The name of the source of the national IDs was available in every case.

**C. Name and address**

18. Every record had a company name. 5300 records did not have address line, 3000 did not have city name. The country and country code were filled by every record.

**D. Legal form**

19. 91 per cent of the records did have legal form.

**E. Activity classification**

20. At least one Nomenclature Générale des Activités Économiques dans les Communautés Européennes (NACE) Rev 1.1 code was available by 91 per cent of the records. We faced a quite high number of invalid NACE codes, 22 per cent of the delivered codes were not valid according to the NACE Rev 1.1 classification.

**F. Employment, turnover, net worth**

21. D&B delivered an estimated variable for employment for every record. Turnover and net worth estimations were also available for every record.

**G. Links between records**

22. Every record had an immediate parent link with DUNS number. 89 per cent of the links did have a time stamp as well. The domestic and global ultimate units were defined by every record. The relationship data between the records were very clear and consistent; this was a strong side of the D&B database.

Table 1

**Number of not null values by variables in the D&B file**

<b>Variables</b>	<b>Filled records</b>	<b>Filled records (%)</b>
DUNS number	330296	100
Company name	330296	100
Address	324975	98
Country code	330296	100
Legal form	300440	91
NACE Rev 1.1, activity code	308061	93
DNB assigned site employment	330296	100
National ID number	265704	80
Type of national ID number	265703	80

## **VI. COMPLETENESS OF THE BUREAU VAN DIJK DATA**

23. The Entities file provided information on the characteristics of the legal units of the delivered MNEs. This file had 505 560 records and 55 variables.

### **A. Identification number**

24. Every record had a BvD number as identification number, the BvD numbers were unique.

### **B. National identification numbers**

25. Of the total population, 73 per cent had national ID number in the BvD database. The type of the national ID was available in every case. 12 per cent of the records did have a "Dun and Bradstreet (DNB) company number", which is the DUNS number of the records from Dun and Bradstreet. These numbers gave additional possibilities to link the data of the two sources.

### **C. Name and address**

26. Every record had a company name. The street address was available by 69 per cent of the records, 75 per cent did have a city name. 9925 records did not have country code; this is 2 per cent, a relatively high number. The EGR can not use records without country code; these records are excluded from the MNE structures. We found 3 invalid country codes as well.

### **D. Further variables**

27. Of the total records, 62 per cent had NACE Rev 1.1 code. 3 per cent of the codes were invalid according to NACE Rev.1.1 classification. For 39 per cent of the records an unconsolidated figure was available on employment. 62 per cent of the records did have legal form. These records were described by 256 different legal forms.

28. The relationships between legal units were delivered in a separate file called Links. This file provided information on the ownership linkages between legal units of the delivered MNEs. Contrary to the D&B database minority ownership relations as above 10 per cent were included as well. This file originally included 588 510 records, 5 duplicate records were removed so it had 588 505 records and 9 variables.

29. The file gave information on the percentages of the ownership relations for every record. The relationships were described by direct or by indirect percentage values. Data on direct percentages were available in 497 000 cases. If the direct percentage was empty an indirect (total) percentage was always available. Time stamp and source description was available for every record.

30. In some cases the file on linkages from BvD included controversial information. It happened that for the same legal unit two different parent companies were recorded, both with 100 per cent ownership. It seemed that BvD tried to collect every piece of information, e.g. from administrative registers, annual accounts, etc, and in some cases they stored different information

on the same relationship. The data were not every time consistent, but these were very few exceptional cases.

Table 2  
**Number of not null values by variables in the BvD files**

	<b>Filled records</b>	<b>Filled records (%)</b>
<b>Variables of the Entities file</b>		
BVD ID number	505560	100
National ID number	368343	73
Type of national ID number	368343	73
DNB company number	61745	12
Company name	505560	100
Address	351282	69
City name	378574	75
ISO country code	495635	98
NACE Rev 1.1. activity code	312120	62
Unconsolidated employment	198970	39
Legal form	312864	62
<b>Variables of the Links file</b>		
BVD ID number of the subsidiary	588505	100
BVD ID number of the shareholder	588505	100
Percent, direct ownership	496875	84
Percent, total ownership	366964	62

## **VII. THE NUMBER OF MULTINATIONAL ENTERPRISE GROUPS BY COUNTRY**

31. The D&B input covered 4216 EU and EFTA groups and 784 non-European groups. The United Kingdom (UK) had the highest number of MNEs in this export, totally 946 records. Reason for this was the high number of UK groups predefined by Eurostat and the non country specific selection which provided a high number of UK groups as well. In case of Greece the D&B database had a very poor coverage. D&B could not match any Greek groups to its database and could provide only the data of 2 influential Greek groups. Outside of this the D&B data set had a good coverage for European MNEs.

32. The BvD input covered all European countries; it included totally 4382 MNEs with European group heads. 923 MNEs had non-European group heads and 147 groups had a private person as group head without country code. These private persons will not be processed by the EGR system; their first subsidiary company will be the group head of the MNE. The UK had the highest number of MNEs in this export as well, totally 871 records. The EGR will include resident MNEs for all EU and EFTA countries.

Table 3  
**Number of MNEs by country and by source**

<b>Country</b>	<b>D&amp;B</b>	<b>BvD</b>
Austria	88	98
Belgium	70	104
Bulgaria	12	11
Cyprus	30	27
Czech Republic	50	28
Denmark	86	99
Estonia	12	12
Finland	97	101
France	411	442
Germany	707	683
Greece	2	39
Hungary	21	11
Iceland	16	17
Ireland	68	70
Italy	243	297
Latvia	8	13
Liechtenstein	12	12
Lithuania	12	12
Luxembourg	95	115
Malta	5	3
Netherlands	418	436
Norway	66	70
Poland	46	53
Portugal	52	54
Romania	9	12
Slovakia	15	9
Slovenia	32	21
Spain	165	205
Sweden	191	235
Switzerland	231	222
United Kingdom	946	871
<b>EU+EFTA total</b>	<b>4216</b>	<b>4382</b>
Non EU+EFTA	784	923
Private persons		147
<b>Total</b>	<b>5000</b>	<b>5452</b>

33. The 5000 MNEs of the D&B final delivery had 330 296 legal units. An average size group included 66 legal units. The input covered 14 group structures with more than 1000 legal units. 811 further groups had more than 100 legal units according to D&B. The largest groups were constituted by 1959 legal units in its tree structure. The BvD final delivery file covered 505 560 legal units. Some records were not part of the 5452 MNEs of the starting population; they were linked only by minority relationships to these groups. The biggest groups covered more than 1000 legal units in this data set as well.

34. In most of the cases the EGR received the data of one enterprise group from both sources. This was the initial goal of the two lots of the data acquisition tender; to have the possibility to combine the information of different sources within the EGR system. Only a smaller part of groups were delivered by one single provider.

35. For MNEs from both sources BvD structures usually included more legal units than the D&B structures. The first analysis showed that the additional units mainly were local units or ceased legal units, which were collected from various sources and their maintenance was not every time successful.

## **VIII. RESULTS OF THE FIRST DATA EXCHANGE WITH MEMBER STATES**

36. In January 2009 the legal units of the two sources were linked at record level by automatic processes of the EGR. In February 2009 Eurostat completed the first data exchange with the Member States; the NSIs received all linked legal units resident in their territory. During this action the NSIs examined the data sets; they linked the records of the data providers to their business registers, they provided them proper national identification numbers. The results of the first data exchange gave information on the quality of the national ID numbers in the source data sets.

37. From the D&B data 235 757 European records were sent to EU and EFTA countries to link to the national business registers. Only 4400 records did not have national IDs, this was only 1.8 per cent. From the 31 EU and EFTA Member States 28 countries cooperated in this action. The countries could link 91.2 per cent of the D&B records successfully to their national registers. They accepted further 5 per cent, these records are national records, but because of register thresholds they are not included in the statistical business register. 1.3 per cent of the ID numbers had to be changed, 1.5 per cent was added. Only 1 per cent of the provided records remained unknown. This proves that 99 per cent of the D&B records are real entities; they are stored in the national business registers.

38. From the BvD data 344 821 records were sent to EU and EFTA countries to link to the national business registers. By this dataset 42 000 records did not have national IDs; this was 12.2 per cent. From the 31 EU and EFTA Member States 28 countries cooperated in this action. The countries could link 80.9 per cent of the BvD records successfully to their national registers. They accepted further 3.7 per cent, these records are national records, but because of thresholds they are not included in the statistical business register. 2.5 per cent of the ID numbers had to be changed, 7.6 per cent was added. 5.3 per cent of the provided records remained unknown.

39. The figures show that BvD data source has slightly lower quality in national IDs. On the one hand for the European records originally 12.2 per cent of the ID numbers were missing (D&B 1.8 per cent); on the other hand 5 per cent of the received records were not able to be linked to the national registers (1 per cent by D&B).

40. During the 2009 cycle the characteristics of the records and the relationships between the units are checked and corrected by EU and EFTA Member States as well. This is the task of the second data exchange in the period July-August 2009. This paper is finalized by July 2009; the



results of the second data exchange with further information on the punctuality of the private data will be presented at the Joint UNECE/Eurostat/OECD Seminar on Business Registers in October 2009.

## **IX. SUMMARY OF THE QUALITY STATEMENTS ON PRIVATE SOURCES**

41. The D&B database is based on control relationships without percentage data. Natural person are not included in the group structures, only legal units are covered. The relationship data between the records is very clear and consistent; this is a strong side of the D&B database.

42. D&B provided the final delivery file to Eurostat according to the defined timetable, the format of the delivered file meets the specifications of Eurostat. The final delivery included the MNEs which were agreed during the selection procedure. The data file had a good coverage on identification numbers (only 1 per cent was missing for European records), on names and addresses, on legal forms, on employment, turnover and net worth figures and on links between records. The activity classification has slightly lower quality; there are a quite high number of invalid codes.

43. The coverage for Greece is poor; D&B could provide data only for 2 influential Greek groups. The quality of the national ID numbers is very good.

44. The BvD database includes percentage values for all relationships. Majority relationships and minority ownership relations as above 10 per cent are included as well. Outside the legal units natural persons are covered as well. The identification of natural persons is difficult; it happened that they do not have country codes, only a name and a BvD number.

45. BvD provided the final delivery files to Eurostat according to the defined timetable, the format of the delivered files met the specifications of Eurostat. The final delivery included the MNEs which were agreed during the selection procedure. The data file had a good coverage on names and addresses, on legal forms and on links between records. The availability of the national IDs and of the activity classification was limited (73 per cent and 62 per cent). 2 per cent of the records did not have country code; within these records there were private persons as group heads as well.

46. Data from BvD usually includes more records for the same group than D&B data. The BvD data set includes a larger proportion of local units and ceased units. The national IDs from BvD have lower quality than in the data set of the other source.

-----