



## Economic and Social Council

Distr.: General  
5 August 2011

English only

---

### Economic Commission for Europe

#### Conference of European Statisticians

##### Group of Experts on Business Registers

##### Twelfth session

Paris, 14-15 September 2011

Item 6 of the provisional agenda

##### **The role of business registers in industrializing the production of statistics**

### **The role of the business register in the context of the strategic vision on industrialization**

**Note by Statistics Netherlands**

#### *Summary*

The business register plays a central role in the system of business statistics. Containing the statistically relevant records on enterprise groups and enterprises, it forms the backbone for the business statistics. In this paper a first attempt is described to redefine the role of the business register in the context of the strategic vision of the High-Level Group for strategic developments in Business Architecture in Statistics.

The views expressed in this document are fully consistent with the vision of the High-Level Group for Strategic Directions in Business Architecture in Statistics, endorsed by the Conference of European Statisticians at its 59<sup>th</sup> Plenary Session in June 2011 (ECE/CES/2011/1 and ECE/CES/2011/CRP.1).

## I. Introduction

1. Traditionally little or no external data were available as input for our statistics, resulting in business processes in which much information had to be asked through surveys. Gradually we have been using more and more official registrations as input for our statistics, resulting in renewed methodologies for the production of statistics with less information that has to be asked additionally. In the information society much data will be available, creating needs for new methodologies, changed production processes, and a redefined role of the business register. In a society that is transforming into an information society, where data are everywhere, the key role of the business register is going to change from providing sample frames to the backbone for data integration from a wide variety of sources.

2. By industrializing<sup>1</sup> statistics the role of a statistician will change. They will play a key role in the translation of the demands from society for statistical information, produced in an industrialized production process that has a variety of sources, both official administrations and any other available registrations (social media, internet, professional organizations that collect (statistical) data, company's registrations). This implies that we have to synchronize the 'outside' world with our statistical concepts. Hence the future role of a statistician will change more and more from a specialist in transforming data from input to output into helping a demanding stakeholder to formulate his demand in a way his need is satisfied and from there creating the desired output from available data.

3. The High-Level Group for Strategic Developments in Business Architecture in Statistics has recently presented their Strategic Vision (HLG-BAS (2011), ref. (1)) as a means of alignment for statisticians in the information society. The process challenge in this vision comprises different and better processes and methods tuned to deliver our products at minimal cost with greater flexibility and in cooperation between institutions. This will help us to improve efficiency and effectiveness.

## II. The business register in the system of industrialised statistics

4. In the information society lots of data is available. In addition to our current official registrations, mostly on a national level, much new information, information on global level and new external datasets can become available as input for our statistics. Business registers should maintain relations between statistical and administrative units as well as the relations with the central banks and other organisations which collect information that is useful to produce statistics of relevance. The basic idea in industrialising statistics is to streamline the abundance of different business concepts, information concepts, "statistical how To" and "production how To". An overview is presented in scheme 1.

5. The Generic Statistical Business Process Model (GSBPM) describes statistical processes. It provides a basis for standard terminology on statistical metadata systems and processes (Steven Vale (2008), ref. (2)). It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources. The GSBPM is a flexible tool to describe and define the set of business processes needed to produce official statistics and helps deriving the statistical methods needed. In the Generic Statistical

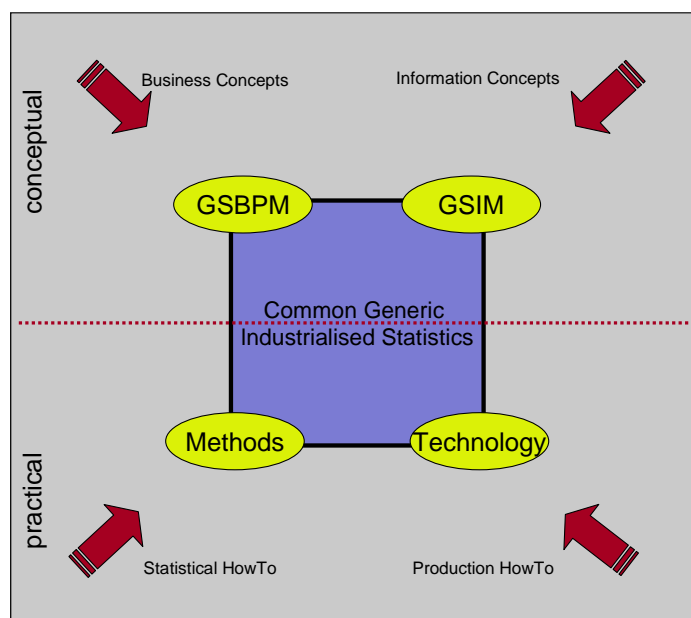
---

<sup>1</sup> Industrialising statistics is the production of statistics based on common and standardised processes, transforming raw data into statistical products according to generic and commonly accepted information concepts (HLG-BAS (2011), ref. (1)-point 19.)

Information Model (GSIM) the information is defined that is required to steer statistical production processes as well as defining the outcome from those processes. The GISM models the objects needed to industrialize statistical processes and describes the relations between them. The GISM forms the basis for the technology needed for the final implementation.

Scheme 1

### Industrialising Statistics



6. The purpose of this paper is to establish the role of the business register in this new way of industrialising statistics according the model described above. Within some examples this role of the business register for the future is explained.

### A. Vacancy statistics

7. One could think of a situation where many enterprises advertise their vacancies over the internet. The challenge is to exhaustively use the vacancy-information from a web-site into a vacancy-repository and link this information direct or indirect to the statistical units in the business register. Then, with some extra methodology, it would not be unthinkable to obtain an internet up-to-date based live register on vacancies. In this case the aim of the business register would not only be to “undouble” the information which is advertised by several web-sites, but also to relate the information on vacancies to standardised statistical information in a coordinated way.

### B. Dutch business architecture industrialising statistics

8. The business register is composed based on official registrations and forms a stable and solid backbone to which data can be linked, including priority rules/quality indicators related to the data and its sources. In this way the Register forms the backbone to which information can be linked (Ritzen (2007), ref. (5)). The data could be processed using a dashboard-like production line, to be controlled and adjusted based on the metadata. In the

next example we will explain the role of the business register in the concepts of the Dutch business architecture industrialising business statistics.

9. Statistics Netherlands (SN) is finalising the redesign of the chain of economic statistics (Huigen (2006), ref. (3)). The derivation of the enterprise group has improved by using extra information about ownershipsrelations known at the tax-office. In the Dutch business register all legal units are divided in two categories: The legal units that belong to the top 1900 largest and complex enterprise groups and the rest. The units in the first group are managed by profiling, where each enterprise group is split “manually” (supported by automatic derivation) into one or more enterprises. For the second group of units we put the enterprise group equal to the enterprise. By this choice, the use of administrative data in small and medium enterprises is optimized, when producing statistics.

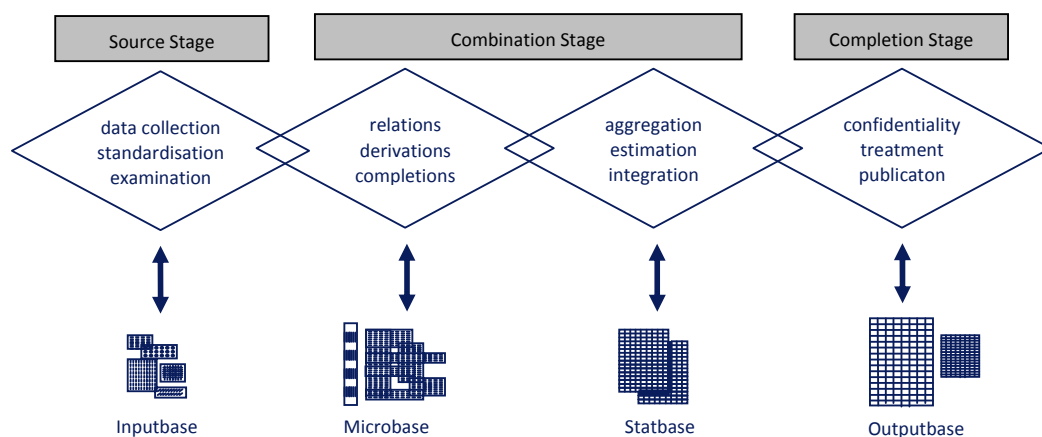
10. Economic data on structure, turnover, investments, innovation, employment or other statistical information, collected by statistical divisions with the largest 300 units, is joined in one central database. Several checks are done on the data in this database by experts (concernprofilers and concerncoordinaters). These experts are accommodated in a separate department within SN and they will correct the observed data if necessary. In this way data on large units is made consistent in order to improve quality and coordinate statistics moreover. The rest of the 1600 other large and complex enterprise groups are managed by profiling. Their data is observed and checked in the isolated survey, but are not viewed in coherence with other surveys.

11. All processes are managed in a coherent manner by chain-management. Strict architectural principles (shown in scheme 2) are used to standardise the business processes of the different economic statistics.

12. Business processes are modelled as coherent sub-processes operating between steady states. The steady states contain data in a well-defined state of processing and of well-defined quality. This facilitates efficient re-use of data. The steady states are grouped into four types of product bases: the inputbase for the raw materials, the microbase and the statbase for the intermediate products and the outputbase for the final products.

Scheme 2

#### Overview of the new Dutch Architecture for the Business Statistics



13. The business register fulfils a special position in the statistical process:

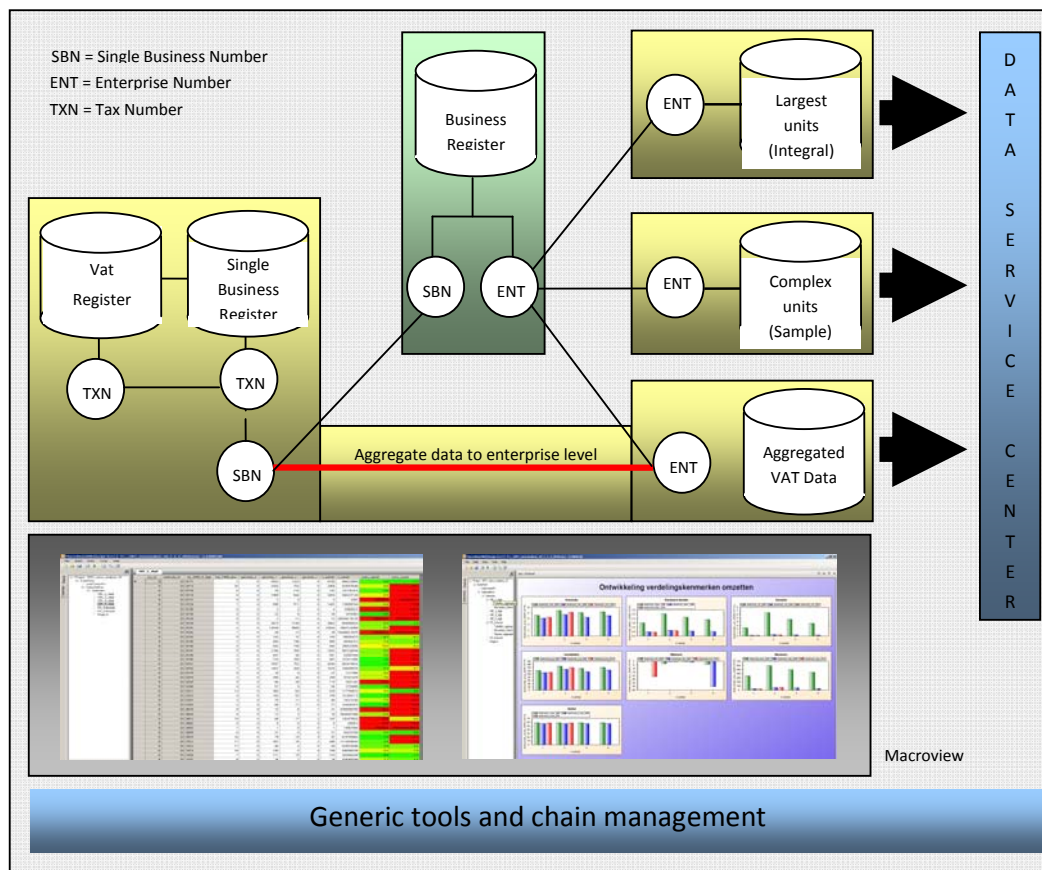
(a) it forms the basis for standardising and linking data from administrative units to the backbone. Aggregation to the level of the statistical unit is used in the combination stage as one flow of input.

(b) it provides the distinction between the largest and complex units on the one side and small and medium sized enterprises on the other side. Within the group of large and complex units, the top 300 largest units are marked extra in the business register. The data of each separated group of units will be used in the combination stage as a different flow of input.

14. Statistical data of small and medium-sized businesses is mainly based on administrative sources in the statistical architecture (as far as the enterprise is not observed in a survey, otherwise survey results are leading). For enterprises that belong to the 300 largest enterprise groups, statistical data are extracted from the central database of largest enterprise groups made consistent. Information about the next 1600 second complex units is gathered in a survey by primary data collection. In the combination stage the different flows of data are combined, aggregated and checked before publication follows in the completion stage.

15. Scheme 3 shows how VAT-data is used for statistics (e.g. growth rates on nace- and sizeclass aggregates) according the model above (Delden & Aelen (2008), ref. (4)). Three (yellow) dataflows are combined in the combination stage and linked to the backbone (green). Each flow is produced by an independent process and results in its own steady state(s). All processes are guided by chain management and guarded by a set of indicators compiled by Macroview. This is a generic tool developed by SN to compile indicators (e.g. dynamics of units, outliers, etc.) for each steady state of data processing with 'uniform looks'. It's possible to show anomalies in several aggregates by different colors and graphics. The final microdata are stored centrally in the data service centre (DSC): a repository of all final statistical microdata and metadata. The DSC is supported by generic tools, which coordinates the exchange of micro- and meta-information.

Scheme 3

**Example of a more industrialised production process**

16. In general, industrialising the production of statistics in the Netherlands starts with a backbone where administrative data is standardised, checked and classified so it can be used to aggregate information from administrative sources to the level of a statistical unit. An important part of the industrialisation process is to build a chain of 'independent' statistical processes with their own steady states. Monitoring the outcome and chain management play an important role to obtain good quality data, as well as the control and the exchange of micro- and metadata is important. Generic tools play a key role in this.

### C. Functional and institutional statistics

17. In general there exists two kinds of Statistics: functional and institutional statistics. Functional statistics describe one specific aspect (function) in society (e.g. external trade, traffic related statistics, surgeries, agricultural activities, fire statistics, etc). Financial or production processes however are often described at an institutional basis, at enterprises or at enterprise groups. A challenge for the future is to link functional and institutional statistics. The business register plays an important role in this process. As an example we discuss the production of statistics on traffic performance of commercial vehicles (lorries, vans, buses) based on odometer readings. At this moment, the approach is functional. Its purpose is to describe the average mileage from the point of view of a vehicle. In the near future, analyses are done on the usage as an institutional statistic. Here we show the value added by the business register for both approaches.

18. Data on odometer readings is available in the database of the National Car Pass, but this information is gathered with other purposes than producing statistics. The Foundation National Car Pass (NAP) prevents and fights against the tampering of odometers and clarifies the odometer before you buy a used car. The coverage of the NAP-database at this moment is about 95%. Approximately 9 million different vehicles are registered by now. Each time a vehicle is visiting an official service station - for example for a vehicle safety check - its odometer-reading is administrated. The registration is updated with 20 million new readings each year. Nowadays the total number of readings is 85 million, which means an average of 9 readings per vehicle.

19. If we want to use this source for statistical goals, then some preparations must be made first:

(a) Define the population of vehicles (vehicle register) and link the usable odometer readings from the NCP-database to the vehicles;

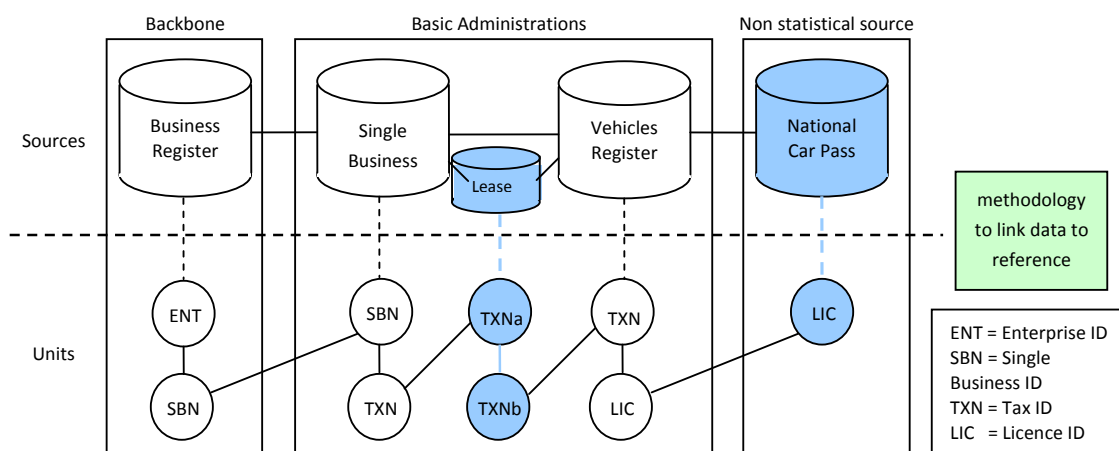
(b) Estimate the mileage for one specific reference year for each vehicle in the population based on odometer readings (methodology);

20. If these issues are clarified in advance, then the total traffic performance in terms of mileage for all vehicles for one specific reference year could be calculated. This aggregate could be broken down by all characteristics of a vehicle (brand, colour, kind of gasoline, etc.).

21. If the tax number of the user of a commercial vehicle is linked to an enterprise in the business register, then for example the nace-code of an enterprise is assigned. The obtained information is used to describe the possession and the usage of a certain type of vehicle in a nace aggregate (functional). Furthermore, if the depreciation cost of the purchase value for each kind vehicle used by a company is added to the total fuel costs per year, then we would be able to compare some aspects on the costs of transport with indicators such as turnover or profit by nace aggregate (institutional). In this way the business register is used as a basic reference frame to administrative registrations from which data on individual units is collected. It fulfils the role of a backbone for the coordination of business statistics in a functional and an institutional approach.

Scheme 4

#### Linking functional and institutional statistics



22. In the vehicle register the tax number of the owner of a vehicle is characterised by the tax number (or the national ID) of the company. This means automatically that all lease vehicles will be related to a lease company. If we want to describe the use of a vehicle, we have to prevent this. Therefore a source is used which holds the relation between the tax

number of a lease company (TXNa) and the tax number of the enterprise that leased the vehicles (TXNb). It is the challenge to add traffic data as a characteristic of a vehicle and then aggregate the information to the level of the enterprise in order to link functional and institutional statistics.

23. One can also think of an extension by linking the licenses of private car users to the general basic registration of all habitants. In this way for example, the possession and the use of passenger cars by sex or age could be compiled. Other possible extensions for the future could be the inclusion of data obtained from navigation-systems or license recognition systems such as average speed or trip characteristics.

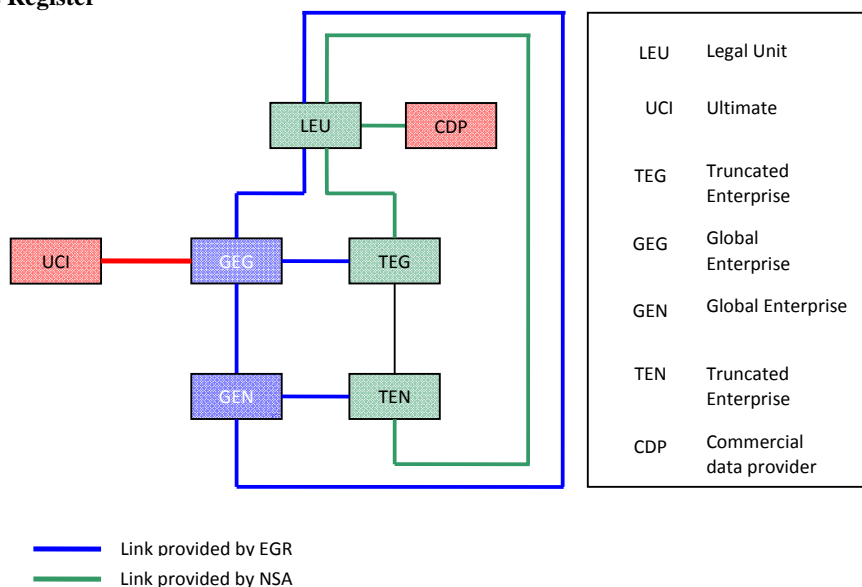
24. In general we can conclude that this approach could also be used as an example to link all kind of functional statistics to institutional statistics. In this respect it is important that a business register maintains the direct or indirect relations between the units used in administrations.

## D. Globalisation

25. Another aim of a business register is to coordinate global enterprises. Each National Statistical Agency (NSA) keeps track of the cross-border relations of units resident in their own country in their own business register. The EuroGroupsRegister (EGR) coordinates all cross-border relations from the member states of the EU and the EFTA-countries. Basic sources of the EGR are originated from a NSA or from a commercial data provider (CDP), such as Dunn & Bradstreet or Bureau van Dijk. The register is a repository with links between the different identification numbers (national identification numbers, identity numbers used in national (statistical) business registers as well as numbers of private data sources, e.g. DUNS number, ISIN number etc.).

Scheme 5

**Eurogroups Register**



26. In the EGR preparations are made to describe globalisation effects by the inward and outward Foreign Affiliate Trade Statistics (FATS). The Inward FATS Statistics describe the domestic population based on those enterprises which are foreign controlled. From a country's point of view the target population consists of all enterprises with an UCI which



is not equal to the country itself. The Outward FATS Statistics describe the foreign population based on those enterprises which are domestic controlled. From a country's point of view the target population consists of all enterprises with an UCI which is equal to the country itself. The EGR also keeps track of 'minority'-cross-border relations on behalf of the statistics on Foreign Direct Investments (FDI).

### **III. Concluding remarks**

27. New dimensions have become important in statistics. Globalisation and administrative data and other upcoming datasources that can be used for statistical purposes in order to minimise survey burden at enterprises have their impact on how the statistical system should function. The development of the use of such data has as a consequence that the statistical systems have become part of the total information system within the national and international society (Ritzen (2007), ref. (5)).

28. The business register can be seen as an important component in opening up and connecting the datasources. It facilitates the new industrialised and standardised production of statistics, both on a national and international level. The business register has evolved into the backbone in the statistical system.

29. Further development of the strategic vision on the new production process is necessary. To help stakeholders in this development to get a clear perspective, we have to continue sharing examples and ideas on possible components of this new production system.

### **IV. References:**

HLG-BAS, High-level Group for Strategic Developments in Business Architecture in Statistics (2011), "Strategic vision of the High-level group for strategic developments in business architecture in statistics", presented at the fifty-ninth plenary session of the Conference of European Statisticians, Geneva, June 2011;

Steven Vale, UNECE-secretariat, GSBPM, General Statistical Business Process Model (2009), version 4.0;

Huigen, R., Project Group Architecture (2006), ICT Masterplan; Statistics Netherlands Architecture, Business and Information model, version 1.0;

Delden & Aelen (2008), Redesigning the chain of economic statistics at Statistics Netherlands: STS-Statistics as an example;

Jean Ritzen (2007), Statistical Business Register, Content, Place and Role in Economic Statistics, paper presented as an introductory lecture on business registers at the ICES-III, Montreal, June 2007.