**Statistics Canada International
Symposium Series - Proceedings**

Symposium 2005 : Methodological Challenges for Future Information needs
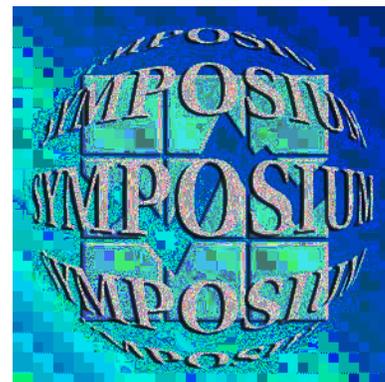
2005

**Statistics
Canada** **Statistique
Canada**

Canadä

THE REDESIGN OF THE DUTCH BUSINESS REGISTER

Kees Vennix[1]

ABSTRACT

The redesign of the Dutch Business Register was started for both technical and statistical reasons. The major changes in the new register are the use of the new Dutch Basic Business Register as the source for legal and local units, the inclusion of administrative units in the register and a new automated algorithm to derive the statistical frame from administrative sources. The new business register will allow a more efficient use of administrative data in the compilation of statistics thanks to an improved linking between administrative and statistical units. In the course of 2006 the new register will be taken into production.

KEY WORDS: Business Register; Redesign; Use of Administrative Sources.

1.   INTRODUCTION

The Business Register of Statistics Netherlands plays a key role in the compilation of most business statistics. It has been in operation since 1967. The current technical system dates from 1993. For a number of reasons it was decided to redesign and rebuild the system. The redesign refers to the process to derive statistical units from administrative information as well as the strategy to maintain the register. Rebuilding the register not only involves the changes caused by the redesign of the process but also those resulting from the transition from one database platform to an other. In practice no software from the current register will be used in the new register, i.e. all functionalities in the new register will be completely new.

The project to redesign the Business Register started in October 2003. It is one of a number of redesign projects that have been undertaken within Statistics Netherlands in recent years. Other projects for example dealt with the redesign of the processes to compile the annual structural business statistics and the introduction of a business architecture. All projects contribute to the purpose that Statistics Netherlands can compile its statistics in a more efficient way: at lower costs, with less response burden for businesses and in better agreement with the changing demands of the users of the statistics. The business architecture project yielded an overall view on the different systems and processes that are necessary to produce statistics. The redesign of the business register has been done in such a way that the new register is correctly embedded in the business architecture.

In Section 2 of this paper the functions of the Dutch Business Register and its contents will be described. In Section 3 the reasons to redesign it will be presented. The major changes will be discussed in Section 4. Subsequently the new process to derive the statistical framework from the administrative sources will be described. In Section 6 the implementation of the new register will be discussed. To finish some remaining problems and future developments will be described.

2.   THE DUTCH BUSINESS REGISTER

Within Statistics Netherlands the Business Register has three important functions. The first one is to supply the population frame for nearly all business surveys conducted by Statistics Netherlands (the main exception is the

_____

[1] Kees Vennix, Statistics Netherlands, Division of Business Statistics, PO Box 4481, 6401 CZ Heerlen, The Netherlands (kvnx@cbs.nl). The views expressed in this paper are those of the author and do not necessarily correspond to the policy of Statistics Netherlands.

international trade statistics). As such it serves as an instrument for co-ordination between different statistics. The second function is to supply name and address information to the survey departments. Finally the business register serves as a bridge between the administrative world and the statistical world. Administrative data become ever more important for Statistics Netherlands in the process to compile statistics. Administrative data become available from different sources and for all kinds of different units. The first step to make these data useful for business statistics is to link them to units in the business register.

The current Dutch business register contains three basic units: the legal unit, the enterprise, and the enterprise group. All units obey the definitions described in Eurostat-regulations (EC 2003). The data model is clarified with an example in Figure 1. The legal unit (LU) is the smallest unit in the register. There are about 1.6 million LU's in the current register. The main source for LU's in the current register is the company register, maintained by the Chambers of Commerce. Because not all economic actors are obliged to register in the company register, also registrations related to social security regulations are used as a source for legal units. This is the case for non-commercial sectors of the economy, for example governmental organisations, public services, education, and health care, and for self-employed professionals like medical doctors and lawyers.

The main statistical unit is the enterprise (ENT). It is defined as the smallest combination of legal units that is an organizational unit, producing goods or services, which benefits of a certain degree of autonomy in decision-making, especially for the allocation of its current resources (EC 2003). In general, an ENT corresponds either to a legal unit or to a combination of legal units. In some cases, especially for public institutions and a few very large EG's, a legal unit is part of more than one ENT. The current register contains 1.05 million ENT. The ENT is the statistical unit that is used for almost all business surveys, including the annual structural business surveys and the surveys for short term statistics. Its main characteristics are the economic activity, classified according to the Dutch, more detailed, version of the European NACE, and the size class, expressed in terms of employment.

The third unit, the Enterprise Group (EG), is defined as an association of enterprises, bound together by legal and/or financial links. In comparison to the ENT, which is autonomous with regard to the allocation of its current resources, the EG is an actor at a more strategic level, taking strategic decisions on behalf of and affecting all of its constituent ENT (EC 2003). The current register contains 0.95 million EG's. They are formed using control relationships between legal units, since an EG can also be considered as an association of LU's instead of ENT. The company register supplies the 100% control relationships between LU's. These are supplemented with information obtained from the EG's itself on relationships involving less than 100% ownership. For multinational EG only the part consisting of LU's that are resident within in the Netherlands is registered.

```
                    ┌──────┐
                    │  EG  │
                    └──────┘
                        │
                    ┌──────┐
                    │ ENT  │
                    └──────┘
                        │
        ┌───────────────┼───────────────┐
    ┌───────┐       ┌───────┐       ┌───────┐
    │ LU 1  │       │ LU 2  │       │ LU 3  │
    └───────┘       └───────┘       └───────┘
```
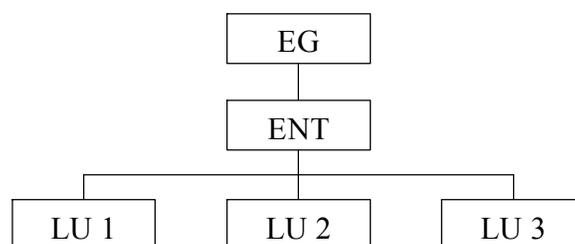
Figure 1: example of the relations between the basic units enterprise group, enterprise and legal unit in data model of the Dutch Business Register for an EG consisting of one ENT and three LU's.

The main characteristics for every type of unit are the economic activity, the size in terms of employment, the address information and the demographic data such as the date of birth and the date of death. In general administrative sources are leading in maintaining these characteristics, but also information from surveys is used in cases where this information indicates that the administrative information was not correct. The number of persons employed is based on information from the social security registrations for legal units with employees, supplemented with estimates for the number of self-employed persons for legal units without employees, and on information from a survey on employment. For all data on employment there is a time lag of roughly one and a half year before the information is registered in the business register.

Most EG (90%) consist of only one ENT and only one LU[2]. For these EG the maintenance of the EG-structure and the formation of ENT are straightforward and done automatically. For larger EG's the formation of ENT's is not always easy. Therefore it is done by hand in the current register, which means that a desk profiler is involved in grouping the LU's into ENT's. For the 5,000 largest EG's the structure in terms of ENT and LU and the characteristics of the units are maintained by desk profilers. This type of maintenance is called active profiling. It means that every one to five years, depending on the size and the economic relevance of the EG, the information in the register is updated and checked with the EG itself. Some 30,000 smaller EG's are subjected to reactive maintenance, i.e. that register data are updated if there is a reason for it from an internal or an external source.

The current register has a production cycle of one month. A month is the frequency of the most frequent short term statistics and it is the frequency at which the data from the company register become available.

3. WHY REDESIGN?

There are technical, statistical and managerial reasons for a redesign of the Dutch business register. The current software and database date from the early nineties. In the course of the years it has become ever more difficult to maintain the system and to adapt to changes required by the users of the registers. Moreover the current process is not transparent throughout and the results are not always reproducible. Although these technical reasons would have been enough to consider a redesign, they were not the main trigger to launch the redesign project. As was mentioned in the previous section, one of the functions of the business register is to serve as a bridge between the administrative world and the statistical world. This bridge function is becoming more and more important, since Statistics Netherlands is aiming to decrease the response burden on businesses by replacing data from surveys by data from administrative sources in the compilation of business statistics. To make this possible it was necessary to improve the link between administrative units and statistical units, in particular the enterprise. Because this required both the introduction of administrative units within the register and a new automated algorithm to derive enterprises, it was considered to be unfeasible within the boundaries of the current register and hence a redesign was necessary to achieve an improved linking between administrative and statistical units.

The triggers to redesign the business register did not only come from within Statistics Netherlands. An important external trigger was the forthcoming introduction of the Basic Business Register (BBR) in the Netherlands. The BBR was started off by the Dutch government as one of a limited number of basic registrations within The Netherlands. The ambition of the BBR is to contain identifying data on all Dutch businesses, thus introducing a unified identification to be used in all registers. This unified identification facilitates the exchange of data between governmental institutions, but it also offers a backbone for the linking of administrative data from different sources. For the enterprises the advantage of the BBR will be that in the long term they have to register only once instead of a separate registration in each register maintained by a governmental institution. It is obvious that the BBR offered Statistics Netherlands the possibility to substantially improve the linking between its own business register and the administrative registers maintained by other governmental institutions.

The first version of the BBR, BBR 1.0, has been in operation since January 2005. It is fed with legal units from two sources: the company register (the main source for the current business register) and the tax office register. To avoid double registrations the legal units offered by these two sources have been linked to each other before being entered into the BBR. In the BBR they receive a unique identification that is also included in the registers which supply the legal units. In BBR 1.0 the characteristics are limited to standardized address information and the legal form of the legal unit. In future versions more characteristics will be available. The maintenance of the BBR is event driven. Every day the changes in the company register and the tax office register are sent to the BBR, in which they are immediately processed.

---

[2] Roughly 0.5 million LU's are not a part of an ENT, because they are considered to be not economically relevant.

4. THE MAJOR CHANGES

The most important changes connected with the new business register are the use of the BBR as the leading source for legal units and the introduction of a fully automated process to derive statistical units from administrative sources. However, these are not the only improvements in comparison to the current register.

In general the new business register will be more flexible than the current one. It will be easier to adapt to changes in the administrative sources and to changes in the needs of the users of the register. On the input side of the register this flexibility is facilitated by making a clear distinction between administrative data and statistical data. This is realized by the introduction of administrative units and corresponding characteristics in the register. An administrative unit is defined as a unit that is registered by an external administrative source. The separation between administrative and statistical units makes it easier to accommodate a change in an administrative source. Such a change will lead to a change in the (process to derive) the administrative units in the business register, but not necessarily to a change in the statistical units or the process to derive those.

Flexibility is also achieved by applying adjustable derivation rules throughout the process to derive statistical units from administrative sources. Rules are adjustable in the sense that parameters within a rule can easily be changed, for example a percentage to be used in some calculation, but also in the sense that different sources can be used for a certain characteristic of a unit in the register. In the cases where more sources for a characteristic are available, these are ranked by priority, accompanied by a time lag. The time lag indicates how long the value of a characteristic from a source is immune for changes in the value of that characteristic from a lower ranked source. Both priority and time lag are adjustable. More flexibility is also achieved in the maintenance of metadata, for example the maintenance of the classification for economic activity. In view of the upcoming NACE-revision (expected in 2008) this is a welcome improvement compared to the current register.

As is the case in the current register, the legal unit, the enterprise and the enterprise group will be the main units in the new register. The definitions and the concepts of these units essentially do not change, but the way they are derived is not the same as in the current register. Moreover, two new units are introduced: the establishment and the local unit. The establishment is defined as the part of an LU that is located at one site. An LU can comprise more than one establishment. In a similar way the local unit is defined as the part of an ENT that is located at one geographical location, defined by the postal code. An ENT thus can comprise more than one local unit. In the current register local units were not included, but constructed on demand, e.g. for regional statistics.

In the new register the company register is replaced by the BBR as the leading source for legal units. The BBR does not only offer the possibility to link to other registers, but it also has a better coverage than the company register. After all, in the BBR the units from the company register are supplemented with those from the tax office register. This offers the possibility to improve the coverage of the business register of Statistics Netherlands in areas where it has traditionally been weak because of the incompleteness of the company register in sectors where registration was not compulsory.

In the process to derive and maintain the new business register information from the tax office will be used in more ways than only indirectly in the form of LU's in the BBR stemming from the tax office register. To begin with, in the new register the number of persons employed for LU's with employees will be based on information on the tax on wages. In comparison to the social security information used in the current register the new source is far more recent (three months versus one and a half year) and expected to be more reliable in the future, because changing legislation and changes in the organisation collecting and supplying the social security data might threaten the quality and the continuity of the current source. For LU's without employees just like in the current register estimates will be used, but the method to make this estimates is more sophisticated than the current one as it uses not only the legal form of the legal unit as input, but also information on the turnover.

As was described in Section 3, in the current register all enterprises consisting of more than one legal unit have to be formed by means of an interactive action by a desk profiler. In the new register an automated algorithm will be used to form enterprise groups and the enterprises within it, regardless of the number of legal units involved. Also in this algorithm information from the tax office register is used (cf. Section 5). The use of an automated algorithm to form enterprises reduces the need for interactive maintenance. The maintenance strategy of the new register will mainly

be driven by changes at the population level, whereas the current strategy is focussed on changes at micro level, i.e. changes in statistical units. Only the 1,000 largest EG's will be subject to active maintenance. For all other units maintenance will be taken up only if changes at population level give cause to do so or if there is an urgent need signalled by a survey department.

Finally a major change will be applied in the way changes in the population frame are registered. In the current register only the birth and death of simple enterprises (EG = ENT = 1 LU) are processed automatically. In the new register statistical events will be used to account for the changes in the frame. These events are derived and classified (birth, merger, take over, …) by an automated algorithm. This algorithm has been described before (Beuken, 2003) and will not be included in the present paper.

5. THE NEW PROCESS

In this section the process to come from the administrative sources to the statistical framework to be used in the business surveys will be described. A part of this process is elucidated in Figure 2.
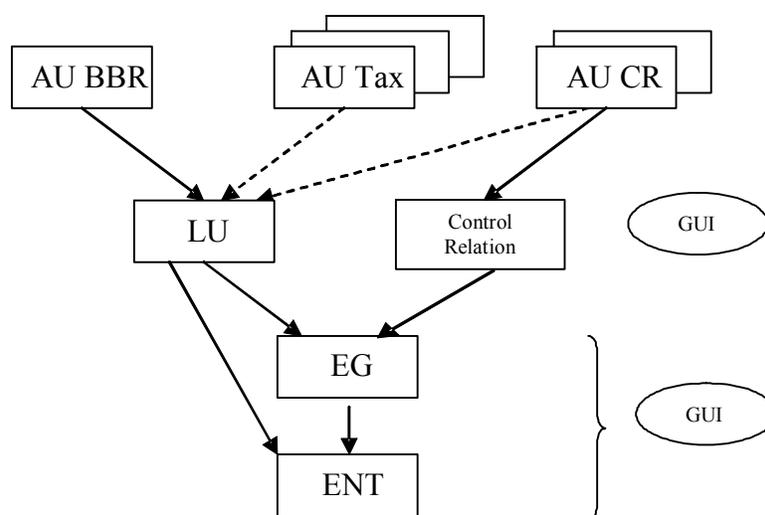
Figure 2: *the new process to derive statistical units from administrative sources. AU CR are AU's based on the company register, AU Tax are AU's based on the tax office register. GUI stands for a graphical user interface by which profilers can apply interactive changes to LU, EG and ENT.*

The first step is to process the information from the administrative sources into administrative units (AU) and their characteristics. In doing so, the information is standardized, checked for irregularities, recoded to the desired classification if necessary, and checked for changes in comparison to the previous state: are there new units, have units been deleted or have units changed? The result is an updated state of the AU's. There are three types of AU's in the business register:

- the AU Basic Business Register: each legal unit delivered by the BBR is transformed into an AU BBR with the characteristics as delivered by the BBR. Since the BBR delivers daily update information to Statistics Netherlands, the AU's BBR are updated every day.
- the AU's based on the Company Register: the Company Register delivers legal units and the 100% control relations between legal units. Both are transformed into AU's. One of the characteristics of the AU for the units is the identification number from the BBR.
- the AU's based on the tax office register: the register of the tax office is quite complicated and comprises several units and several types of relationships between those units. The main ones for the use in the business register at Statistics Netherlands are the "person" (the unit that carries the BBR-identification number, comparable to the legal unit) and the three composite units, consisting of one or

more persons, by which a business fulfils its tax obligations. These are the units for the Value Added Tax, for the corporate tax and for the tax on wages. Each unit has its own characteristics. In principle all are linked to one or more BBR-units by the BBR-identification number registered on the person(s) constituting the AU, but due to deficiencies in the linking between the tax office register and the BBR some AU cannot be linked to a BBR-unit.

After AU's are formed, the next step is to register the legal unit and its characteristics. As was pointed out in the previous section, the BBR is the leading and in fact the only source for LU's, which means that each AU BBR becomes an LU. Every LU is linked to:

- 1 AU BBR
- 0 or 1 AU corresponding to a legal unit in the company register
- 0 to n AU corresponding to a 100% control relationship from the company register
- 0 or 1 AU corresponding to a person in the tax office register
- 0 to n of the composite AU based on the tax office register.

The current version of the BBR contains only a limited set of characteristics. The required characteristics for the LU that cannot be taken from the BBR are derived from one of the AU's linked to the LU. For each characteristic there is a derivation rule that describes which source (AU) has the priority. Once the LU and its characteristics have been registered, its economic relevance is checked to prevent marginal LU of becoming an EG or an ENT. Two indicators are derived: one that indicates whether or not the LU on itself should lead to an EG, regardless of the control relations it might have with other LU's, and one that indicates whether it should lead to an ENT. In the algorithm to derive these indicators the number of persons employed, the NACE-code (auxiliary or not) and the turnover of the corresponding AU for the value added tax play a part.

After the LU is registered the enterprise groups are formed. In the formation of the EG the 100% control relationships from the company register are used, linked to the LU by the link to the corresponding AU, and the control relationships involving less than 100% control that are registered by the desk profilers during their maintenance of EG's (there is no administrative source for these less than 100% control relations). Also control relationships are subjected to derivation rules with priorities and time lags.

Once the EG is formed the enterprises within it have to be formed. As was mentioned in Section 4, an automated algorithm has been implemented in the new register for this purpose. For by far the most EG it's simple: the EG consists of only one ENT and only one LU. Therefore the first step in the algorithm is to check for the number of LU's in the EG. If there are two or more LU's in the EG the remainder of the algorithm is followed. That remainder consists of 26 branches, involving the following aspects:

- the number of LU in the EG: two or more than two?
- the activities of the LU's: are there auxiliary activities involved or not?
- the number of persons employed: is it zero or larger than zero?
- do the LU's involved share a common pay roll?
- are the LU's part of the same corporate tax unit?

In answering the latter two questions the AU's based on the tax office register are used. In practice the algorithm first tries to group all LU's within the EG into one ENT. If this turns out to be impossible, the same algorithm is applied to try to group all LU's that are in the same corporate tax unit into one ENT. All LU's that are left after this step and that are considered being economically relevant become an ENT on their own.

By linking the legal units to the AU based on the tax office register an improved linking between the business register and the most important source for administrative data that might replace survey data is already established. However, surveys are not held among LU's, but among enterprises. Therefore better linking between administrative units and the ENT is of more importance to facilitate the use of administrative data in the compilation of business statistics. In a prototype for the new register it turned out this could be achieved by using the AU for the corporate tax in the formation of ENT. Not only the disclosure of corporate tax data improved significantly, but also that of the

value added tax information. This offers the possibility to improve the integration of data from the international trade statistics, which is based on value added tax units, with those from other business statistics.

After the EG and ENT are derived two steps of the automated process are left: to derive the preliminary statistical events and to check for continuity. The events indicate what changes happened to a statistical unit since the last time the statistical frame was authorized (see below). For those units that changed, it is checked whether the unit can be considered as the continuation of an existing unit, using an algorithm based on retention of 70% of the number of persons employed. If so, the identification of that unit is retained.

The process to derive EG and ENT from administrative sources is fully automated. Every night the information from the administrative sources is processed to deliver a preliminary frame. During the day the desk profilers can interactively apply changes to this frame in case they do not agree with the result of the automated process or if they have information about a unit from an other source, for example from a survey or from a contact with an EG, that needs to be registered in the business register. Changes can be applied to several things:

- control relationships between LU's (both the LU's involved in a control relationship and the type and percentage of control can be changed)
- the composition of EG and ENT
- characteristics of EG, ENT and LU
- the decision of the system regarding the continuation of an ENT or EG can be overruled

By verifying that predefined business rules are obeyed by the changes applied interactively it is assured that the database retains its consistency. For the Top 1000 EG's that are the subject of active maintenance by profilers the changes suggested by the automated process are presented to the profiler before they are processed in the database. The profiler can accept or reject the changes or modify them in such a way that they yield the desired result.

Just like the current register the new register will have a production cycle of one month. Every day a preliminary frame is produced as the result of the changes in the daily updated BBR and the changes applied by interactive profiling. Every day the state of the frame is checked by macro editing, which in first instance is aimed at changes at population level that occurred since the last time the frame was authorized (see below). The system produces overviews of the changes per size class, per NACE-code at different levels and per combination of size class and NACE-code. From these overviews the macro editor can trace suspicious areas in the register which deserve closer attention. In practice the macro editor can select the statistical events leading to the suspicious changes and assign these to a profiler to verify whether these events are correct.

Once a month the preliminary frame is authorized and stored in a separate environment where it is accessible for the users of the business register. When the frame is authorized, the preliminary statistical events accounting for the changes since the last time the frame was authorized become authorized events. The authorized frame is the statistical frame that is used for all co-ordinated business surveys of Statistics Netherlands. Because of the strict demands regarding the reproducibility of this authorized statistical frame and the effort that is needed to guarantee that reproducibility not the whole business register is stored in the separate environment. The authorized statistical frame contains:

- the enterprise groups, the enterprises, the legal units and the local units
- the links between these units
- the characteristics of units and links as far as they are needed for identification and stratification
- the statistical events accounting for the changes since the previous authorized frame

The reproducibility of the frame is guaranteed by two mechanisms: the statistical events and the inclusion of three dates on the units, the links and the characteristics that are part of the authorized statistical frame. The statistical events make the statistical frame reproducible because by registering which events lead from a certain state of the statistical frame to the next one and vice versa the state of the statistical frame at any given moment can be reproduced. The three dates that are stored in the authorized frame concern the date, at which a change occurred in reality, the date at which the change was registered in the business register and the data from when on the change was applied in the statistical frame. Combining the latter two dates in the proper way when making a selection from

the statistical frame allows to reproduce the state of the frame as it was known at any given date in the past. The first date can be used as an indicator for the quality of the register.

Note that address information is not included in the statistical frame. The new business register will strictly follow the administrative sources with regard to address information. These standard addresses will be supplied to the survey departments. If survey observations yield deviant address information, which will often be survey specific, it will be registered in the central contact and observation register system of the business statistics division of Statistics Netherlands and not in the business register. This is in line with the business architecture.

6. IMPLEMENTATION OF THE REDESIGN

The redesign project started in October 2003 with an inception phase in which the scope of the project was established in consultation with the register department and the users of the business register. This resulted in a list of needs to be fulfilled by and features to be included in the new register. The year 2004 was dedicated to the elaboration phase. The new algorithms for the automated derivation of enterprises and statistical events were developed and tested in practice by the construction of prototypes of the software. These prototypes turned out to be very useful to detect where the algorithms needed to be improved. At the same time they yielded the confidence that the choices made while designing the new concepts during the inception and elaboration phases were correct, realistic and feasible. In October 2004 the construction phase started in which the eventual software and database of the new business register are actually being constructed. Originally the aim was to take the new business register into production by January 2006. Unfortunately this turned out to be too optimistic and hence the implementation had to be delayed until a date in the course of 2006, most probably July 1.

One of the steps towards the implementation of the new register is the transfer of data from the current register to the new one. It is accepted by the users of the business register that changes to units and their characteristics are inevitable, considering that new sources for the legal units and for the number of persons employed are used and that a new algorithm to form enterprises is introduced. However, the current register contains information that cannot be overruled by the results of the new process because they are based on agreements with and information from the respondents. Most notable examples are the composition of ENT's for the largest EG's and the characteristics used for stratification. If a value from the administrative source is replaced by one based on information from the ENT itself in the current register, this information is retained in the new register if the ENT returns with the same composition. For some 35,000 "protected" ENT's belonging to the 2,500 largest EG's in the current register the composition of the ENT is retained in the new register as far as possible.

After consulting the users of the register it was decided to have six months of history according to the new concepts in the new register at the moment that it will be taken into production. This is achieved by deriving the new register from scratch step by step according to the process described in Section 5. After each step the result is compared to the current register. For the units where information from the current register is to be retained, this information is processed in the new register before continuing with the next step. This means for example that after control relations are derived from the administrative sources in the new register, the situation is compared to that in the current register before EG's are formed. If there are control relations for the protected EG's in the current register that are not yet there in the new register, these are taken over in the new register before the EG's are derived. This process will be repeated every month until the new register will be in operation. In fact it comes down to copying the results of profiling activities done on the current register into the new one. Changes that have been applied in a previous month are protected against changes due to the administrative sources by the normal mechanism based on priority and time lag. The last step in each month is to copy the identification numbers from the current register for those units that can be considered as a continuation of a unit from the current register, using the same algorithm that will be used in the new register.

7. FUTURE DEVELOPMENTS

Not all register problems will be solved by the implementation of the new business register. There are some well known issues that will need extra attention from the register department just like in the current register. One of them

is the phenomenon where according to the administrative sources one LU in an EG employs all the persons working in the EG, whereas in practice they work for different LU's and ENT's. The new register offers the possibility to divide the personnel over more than one ENT within the EG, including ENT's of which the LU is not a part of. The problem however is to detect those cases. An other long running problem involves partnerships, a legal construction that is often used in the medical and law sectors. The BBR delivers both the partnership and the partners as LU's, but not the relationship between them. Depending on which LU the personnel and the turnover are registered, the partnership and the partners can all become separate ENT and separate EG. In the future the use of information from the tax office register could help to establish the relations