

# Multi-source data integration with SBR

Authors:

Steen Eiberg Jørgensen, Head of SBR

Ilsabe Børsting, Senior Adviser

Christian Bendtsen, IT-specialist

## 1. INTRODUCTION

Statistics Denmark's (SD) Statistical Business Register (SBR) is very closely linked to an Administrative Business Register (Tax ABR) as well as a Central Administrative Business Register (CBR) administered by the Danish Business Authority. Together, the three institutions own and develop CBR. SD is mentioned as a data partner in the Central Business Register Act. The CBR and SBR are covering units from all institutional sectors.

Please refer to the below Figure 1 for an overview of the sources for the CBR in Denmark, including relations to the Statistical Business Register. Statistics Denmark is online with the central CBR.

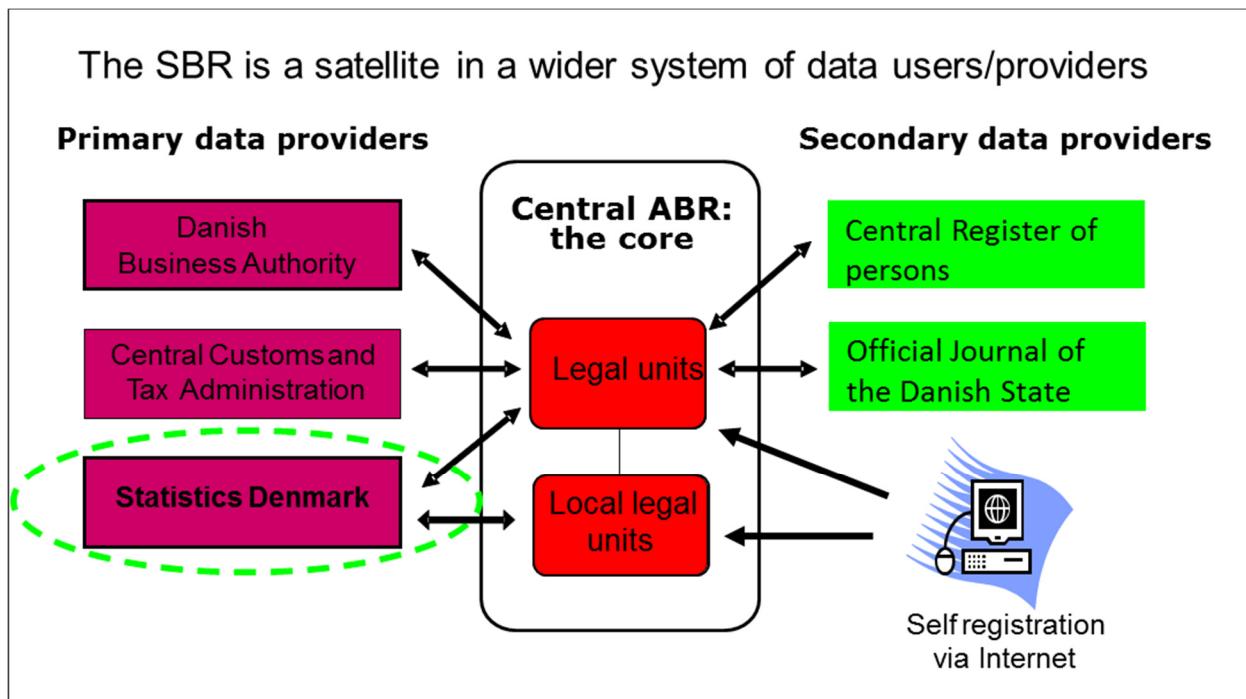


Figure 1. Basis for the Danish SBR system.

In the past few years, CBR has been developed with more information from other business registers that have become digital and often with the enterprise in charge of the update.

Information about employees is received from the Danish tax authorities at least monthly. Companies are obliged to report to the tax authorities no later than the last day of the month for payments in the tax month. The companies report information about employees for legal units as well as local legal units. In addition to payroll information, the report includes personal data, hours worked, and pay period.

## 2. BACKGROUND

Sources for determination of relations are constantly developing. For this reason, SD has decided to look into SBR's loading of owner relations.

In terms of group registration, CBR now contains registration of *beneficial owners* (since January 2018) and *legal owners* (since 2014). In case of changes in ownership, the enterprises now have to update the ownership information within three weeks.

Statistics Denmark downloads information about beneficial as well as legal owners. However, we are only using our knowledge about legal owners at present. We have not made any analysis of the quality of the registration of beneficial owners. It is probably relevant to submit this information to the European Group Register (EGR).

In addition to the known statistical units, SBR also has a Group unit. Groups have been included in our register since 2008. In 2017/2018, the data model for groups has undergone updating to allow us to operate with multiple sources, whereas the previous system only had one Commercial Data Provider (CDP).

The Act on Statistics Denmark stipulates that the development of registers must factor in SD and their needs. SD was indeed heard in this case; unfortunately, we did not succeed in changing the setting of intervals that was decided, where the interval 50-66 per cent ownership/voting rights creates problems for us in the creation of groups. In this case, a limit of >50 per cent would have been preferable, since the control is otherwise not quite clear.

### What are

The law defines a **beneficial owner** as a natural person who ultimately owns or controls a company through direct or indirect ownership or a sufficient percentage (25 per cent) of the shares or voting rights, or through control via other means.

**Legal owners** are persons and corporate entities having direct ownership or voting rights in a company. All persons and corporate entities having direct ownership of five per cent or more of a company's capital or voting rights must be registered.

For further details, see the "Act amending the Companies Act, the Certain Commercial Undertakings Act, the Corporate Funds Act and various other acts" on the Danish Business Authority's webpage.

In the updated version of groups, Statistics Denmark wanted to include information from EGR and updates from FATS to better cater for e.g. the FATS statistics. However, there is a demand for data on groups from producers of statistics as well as from external customers who are asking for statistics on this.

To sum up, the main issue facing the development of the Danish Group Register is the integration of multiple sources. The most important challenges in this regard are determination of control; timeliness; and coverage.

## 3. DATA MODEL

Please see Annex I.

In building the data model, the aim has been to ensure the maximum amount of automatic processes in the treatment of Danish MNEs. For this purpose, certain quality programmes will be applied. Please refer to the data model in Annex I for an overview of troubleshooting processes, eg. priority rules, rules for overriding data etc.

## 4. GROUPS TREATED BY THE LARGE CASE UNIT IN SD WILL BE HANDLED MANUALLY. THE SOURCES

Our commercial data providers (**CDP**) continue to be our primary source. They rest primarily on data from CBR, but

they also handle units for which the data quality in CBR is not good. We pay them to continuously quality assure data and we receive a monthly delivery, which we load continuously.

**EGR** is a source that we use to load relations containing Danish entities. EGR will be a secondary source for relations, where an EU country is one party in the relation and the relation is to a subsidiary of a Danish entity.

We do not use **Annual reports** as a primary source in any way. We use annual reports in connection with **Profiling** and they are an important element in understanding the enterprise and in checking whether we have registered all entities of the group.

Troubleshooting and reporting to I-FATS and O-FATS will provide input for what we call Statistics Denmark corrections.

## 5. METHODS

When we weigh the sources on one another, it is important to pay attention to when we received the information and from when it is valid.

Below please find an example of data reception from our CDP and EGR, respectively, with an owner relation between a German and a Danish enterprise.

### Example 1

#### From CDP

DE-LeU A owns DK-LeU A from 1-5-2017 received 1-4-2018

#### From EGR

DE-LeU A owns DK-LeU A from 1-1-2017 received 1-8-2018

Based on accounting year for the Danish enterprise, we will prioritise from when the change must be valid.Example 2

#### From CDP

DE-LeU B owns DK-LeU A from 1-1-2018 received 1-4-2018

#### From EGR

DE-LeU A owns DK-LeU A from 1-1-2017 received 1-8-2018

We will make sure that both rows exist in our group register

#### *Result in the Danish Group Register:*

DE-LeU A owns DK-LeU A from 1-1-2017 to 31-12-2017 received 1-8-2018

DE-LeU B owns DK-LeU A from 1-1-2018 to xx-xx-xxxx received 1-4-2018

If we are in a situation where we have loaded the data ourselves via troubleshooting, the data will not be overwritten for the year in question. However, if we receive subsequent changes, the data will be updated.

### Example 3

#### From EGR

DK-LeU A owns DE-LeU A from 1-1-2016 received 1-3-2018

#### Own (SD) correction

DK-LeU B owns DE-LeU A from 1-1-2017 received 1-4-2018

#### From CDP (Same data as received 1-3-2018)

DK-LeU A owns DE-LeU A from 1-1-2016 received 1-8-2018 (will not be loaded)

#### From CDP

DK-LeU C owns DE-LeU A from 1-1-2017 received 1-11-2018 (will not be loaded)

It is still our own registration covering the same period that has the highest priority.

#### From EGR

DK-LeU D owns DE-LeU B from 1-1-2018 received 1-3-2019 (will be loaded)

This change will override our own correction, as the “from date” is after our “from date” and we look at it as new information.

The active rows in the Danish Group Register will be as shown below:DK-LeU:

A owns DE-LeU A from 1-1-2016 to 31-12-2016 received 1-3-2018

DK-LeU B owns DE-LeU A from 1-1-2017 to 31-12-2017 received 1-4-2018

DK-LeU D owns DE-LeU B from 1-1-2018 to xx-xx-xxxx received 1-3-2019

## **6. CONCLUSIONS**

A recent act (2016) amending legislation regarding legal and beneficial owners has prompted a major revision of the data model of the Danish Group Register. This was due to our analyses of data quality, which showed the necessity of handling multiple data sources. Looking ahead, there was also a desire to prepare the Danish Group Register for data from EGR.

When operating with multiple sources, it is important to have rules of priority, also with regard to dates. Also, when looking at multinational companies, we find that more than one source is a necessity in order to obtain the full picture. As illustrated in example 3 in the above section 5, multiple sources raise the need for increased quality checks, especially concerning dates.

Our expectation is that especially the from\_date and to\_date will be improved with our own corrections in relation to what we get from CBR and CDP. It is our hope that this will result in a much better overview of control relations and thus the group structure. More importantly, this raise in quality should encourage e.g. FATS users to use the Group Register. Ultimately, results from FATS surveys should feed back to the Danish SBR and possibly EGR.

## ANNEX I: DATA MODEL

