

eIncome – A Unique New Source of Administrative Data and the Impact on the Quality of the Business Register and a Number of Surveys

Paper for the Seminar on the “Use of Administrative Data in Business Statistics” - 18-19 March 2010 in Rome

Policy on register-based information in Denmark

It is a characteristic feature of Denmark that everything under the sun is widely registered. It is also stipulated in a Danish Registers Act that “register-based information shall not be misused”. The Data Surveillance Authority is responsible for ensuring that legislation relating to registers is not violated. Citizens and businesses can apply to the Data Surveillance Authority, if they have any complaints against the contents or use of the registers.

Central Population Register (CPR)

Since 1968 Denmark has operated a register of all persons who are living or who have lived in Denmark, called the Central Population Register (CPR). The CPR is updated daily on the basis of the population registers operated by each municipality in which the citizens live. Information is passed on to the municipal population registers by the different authorities that are of relevance. All registered persons are allocated with a civil registration number consisting of 10 digits. The civil registration number is, e.g. used for identifying people, and not two civil registration numbers are alike. Private individuals and businesses may obtain access to some of the information registered in the CPR. However, access can be restricted as required.

It is common practice for the individual citizen to state the civil registration number when contacting a bank, physician, dentist, library, or engaged in a credit sale, etc. Consequently, it is possible to obtain a highly detailed picture of each individual citizen by combining data from various registers.

It is generally unlawful to link registers

By using the CPR number as key opens up the possibility of achieving a detailed picture of each individual citizen. However, legislation relating to registers is generally very restrictive in Denmark, which implies that it is unlawful to link registers containing personal information. This restriction has resulted in difficulties for the social authorities and tax authorities with regard to checking the correctness of the information provided by the citizens, prior to payment of transfer benefits, due to e.g. unemployment, illness.

Act on eIncome

The Act on an income register was adopted in 2007, and subsequently data for the new register were to be reported for payments with effect from 1 January 2008. The Act is administered by the Danish tax authorities.

From the date of when the Act came into effect, all public and private employers as well as public authorities paying out money to the citizens at least once each month must report detailed information on the size and the type of payment. This opens up the possibility for the public authorities to conduct checks of each individual citizen's finances by entering the register, which is thus updated monthly.

In contrast, information is only to be reported once to the authorities, as the data are distributed to the relevant authorities requiring these. It is hereby ensured that the response burden of each individual business is reduced.

Use of data in the new register is subject to checks

As the new register contains a wide range of sensitive personal data, access to the data is therefore very restrictive. By submitting an application, it is possible to obtain access to the data, but the reason must be well-founded in order to be permitted access to each and everyone of the numerous variables in the register.

However, Statistics Denmark has special access to all data, provided that the data are used for statistical purposes only.

The eIR is used as primary data source by Statistics Denmark

New information used for statistics

All data are currently collected by Statistics Denmark (on average, 10 million data reports per month) and are subsequently refined for statistical purposes in Statistics Denmark's own register, called the eIR (electronic Income Register).

Previously, data were by far obtained from the Danish tax authorities, but only once each year. Monthly data reports imply that short-term statistics can be produced and not only structural statistics.

The eIR contains information on each job at a very detailed level, i.e. one single hour's paid work per month per employer. The data reports of the employers also state the production unit (LKAU, workplace) at which the employee is employed. This opens up the possibility of registering all jobs that are held by each individual employee with different employers.

Consistent primary data for many statistics

One of the greatest benefits of the new data source is that it forms the basis for compiling a great variety of statistics, which were previously based on different data sources. The definitions for a number of statistical concepts can subsequently be standardized, and a greater consistency is achieved when comparisons are made across the statistics.

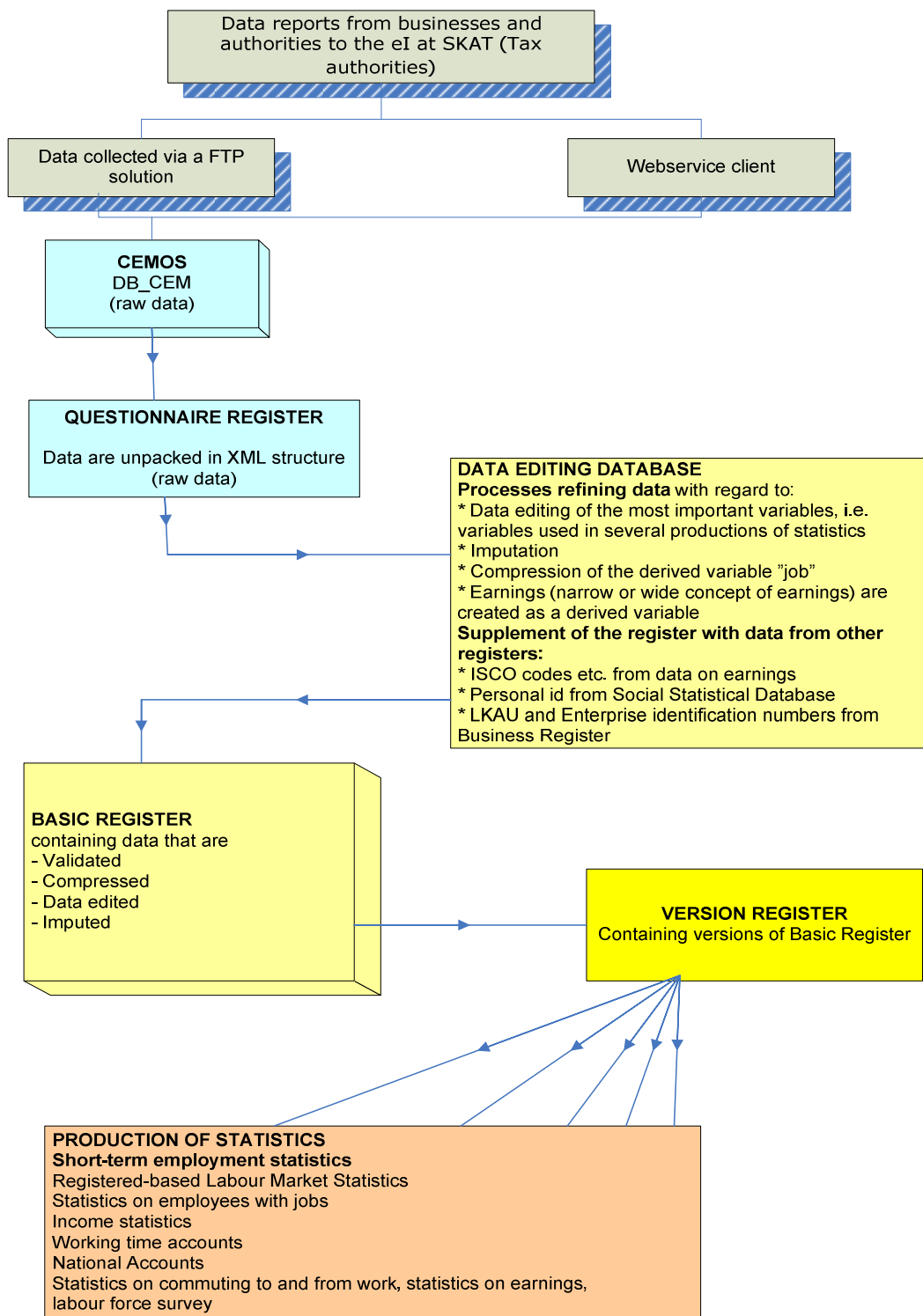
However, it is also a prerequisite of achieving consistent primary data that changes in the primary data due to errors, omissions, etc. are handled in the same way. In the light of this, it is intended that changes in the raw data are handled centrally in the eIR, and that statistics are not compiled on the basis of raw data, but exclusively on the basis of a version of the register, which is "released" for statistical purposes in order to exactly ensure that the statistics are produced on the basis of homogeneous primary data.

Structure of the eIR system

The register was set up by Statistics Denmark for the following purposes:

- The register must be able to handle vast amounts of data, as 10 million records are, on average, received each month.
- All data reports are stored in raw form in the Basic Register. This implies that recreation of the raw data will always be possible.
- The users, who are producing statistics on the basis of the register, obtain access to standardized versions of the register in order to ensure the primary data of the statistics are consistent. Consequently, monthly versions are preserved in a Version Register, the users of statistics are only permitted to obtain access to the data in this part of the eIR system.
- All changes to previous data reports and additions in the form of imputations, supplements from other registers operated by Statistics Denmark, etc. are preserved in a special Data Editing Database.

Schematic diagram of the eIR system



Limitations of the new source

The eIR does not contain information on:

- Income from self-employment – which is the reason why only employment statistics on employees can be compiled and not on total employment
- Incomes and payments from abroad (e.g. income earned by Danish citizens resident in Denmark but working in Sweden or Germany)
- Income from shares and capital income.

Furthermore, the quality of the Register is limited by the quality of the data that we receive. Reporting is compulsory but not according to Statistics Denmark but to the tax authorities. This puts constraints to the possibilities of affecting the data reports directly from businesses.

Data quality in eIR

Statistics Denmark receives raw data containing a number of errors and omissions that need to be corrected in order to be able to produce statistics providing a representative picture. Especially, data on earnings per hour worked are subject to errors as well as the indication of correct LKAU. But also amounts of money, period indications, etc. may be subject to errors, and this implies, consequently, that great demands are placed on the process of data editing.

The tax authorities are primarily concerned about assessment of taxes once every year which is too rarely a frequency for the purpose of Statistics Denmark.

eIR focuses on improving the quality of the variables most important for making short term employment statistics. We have started with validating and improving the quality of:

- Number of jobs
- Number of employees
- Earnings.

By using validity checks also based on input from other sources data reports are compressed, edited and better data are imputed.

Business Register's use of the eIR

The data quality of the number of employed persons at the level of LKAU is improved

Up to now, the information in the Business Register (BR) on the number of employed persons at the level of LKAU has been updated once each year. This has been performed on the basis of the questionnaires that have been sent out and has, subsequently, put a strain on the response burden as well as resulting in additional work for Statistics Denmark. In addition to the annual questionnaire-based updating, the statistics on employees with jobs are quarterly compiled as a short-term indicator on the basis of administrative data. However, the quality of these data has been subject to considerable margins of statistical inaccuracy.

The establishment of the eIR implies that data on amounts paid out by employers to employees with information on the employees' workplace, LKAU, are now received at least once each month. This will lead to a considerable improvement of the BR's information on the number of employed persons distributed among LKAU.

Jobs are registered in the eIR at the most detailed statistical unit, the LKAU and can therefore easily be aggregated at the level of KAU and at the level of Enterprise.

Better quality in the BR results in an improved quality for many statistics

As the number of employed persons per statistical unit in the BR is a relevant background variable for many statistics, an improvement of the variable in the BR will also result in an improved quality for all the statistics using employment as background variable in connection with data extracts from the BR. Consequently, the quality of questionnaire-based survey will be improved and the same is true of the result of the subsequent process of raising the figures.

Imputation of missing LKAU in eIR and BR

Unfortunately a rather high part of data reports do not include information about the LKAU although it is in conflict with existing laws and regulations. In case of missing LKAU a standard procedure ensures imputation of a valid code. The rules are:

- If an identical report exists from last month for the employee the same LKAU is assigned as last month.
- If the data report is new the LKAU is imputed from the employer's LKAU closest to the address of the employee.

The imputed LKAU depend on whether the employment condition is in the private or public sector but the LKAU is roughly validated as regards

- LKAU is registered as active
- the NACE code of the LKAU imputed harmonizes with the ISCO (International Standard Classification of Occupation).

Also the LKAU is validated on the basis of available data from existing databases as e.g.:

- whether the job is in the private sector or public sector
- whether the NACE code of the LKAU imputed harmonizes with the ISCO (International Standard Classification of Occupation).

An effort will be made in near future regarding raising the quality of the LKAU reported by employees, as not only Statistics Denmark but also the tax authorities are interested in knowing the exact LKAU of every employee in order to calculate automatically the travelling expenses which are tax-deductible to a certain extent.

New possibilities of compiling statistics on the basis of the eIR

In addition to general quality improvements of a number of statistics as a consequence of the future updating of employment figures and workplaces in the BR, which is conducted via the eIR, the register will form the basis for compiling a number of new statistics:

New short-term statistics A number of short-term statistics are in the process of being compiled on the basis of the new eIR. The statistics can be published by geographical area, where the location of the workplace is adopted as starting point, but also the place of residence of the employee can be incorporated as background variable. The statistics will be published quarterly, but it should be possible to publish monthly statistics in the long term.

New short-term statistics based on eIR:

- New employment statistics (including wage earners only). The statistics can be compiled by full-time employed persons and by the number of hours worked by each individual wage earner, e.g. each month.
- New job statistics, where the total number of jobs can be compiled as well as the number of jobs that have been held by each individual employee.
- It will be possible to compile the statistics on aggregate wage and salary costs for each job and each employee.

Impact on long-term statistics A number of structural statistics will in future use the eIR as data source, which improves the consistency between the short-term and structural statistics. Simultaneously, it will be possible to use more standardized concepts. This applies to, among others, the following structural statistics:

- Employed Persons
- Salary Earners by sector
- Labour Force Survey
- Working Time Accounts
- Commuting to and from work
- Personal income (exclusive of self-employed).

Longitudinally based data register

A new great potential opened up by the new register is the possibility of developing an overall register, which shows the movement of a person from unemployment to employment over time and conversely.

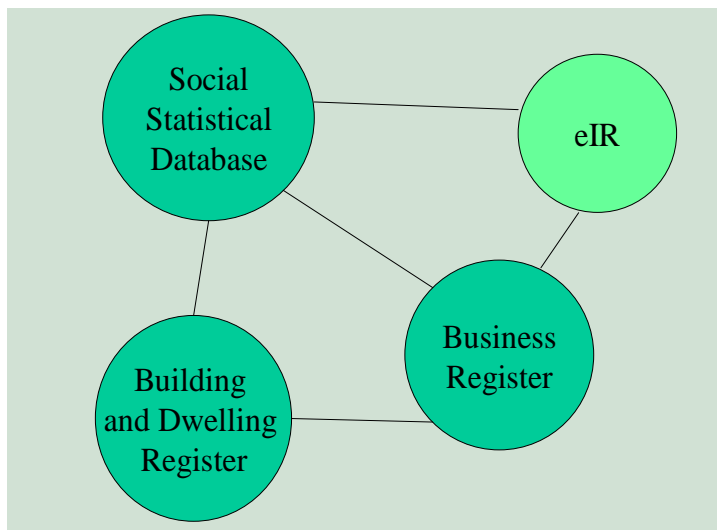
Statistics Denmark already keeps a register, where persons who are not in ordinary employment can be followed over time. The register contains data on unemployed persons claiming unemployment benefits and social assistance, but also persons participating in different types of labour-market financed measures, e.g. flex jobs, traineeships, receiving training and education, etc.

Using the CPR number as key, data can be combined so that it is possible to follow each individual person over time, irrespective of whether this person is unemployed, not in ordinary employment or employed.

The establishment of Statistics Denmark's overall register, which contains information on persons in the labour market, irrespective of whether these persons are employed or not, opens up the possibility of compiling a range of completely new statistics.

Combination of the eIR with other statistical registers

It will generally be possible to combine information from the eIR with information from the other major registers kept by Statistics Denmark, as both the CPR number and the BR units function as keys, which provide access to data from other registers, and thereby almost unlimited possibilities of monitoring different statistical areas.



Using LKAU as key, information on businesses and their workplaces from the BR can be linked with the eIR.

A third and major register, which is widely used in producing statistics, is the *Building and Dwelling Register* established in 1976 and which contains a precise registration of all properties and buildings in Denmark. The information in the BBR reflects the actual, physical conditions of the housing stock. By means of addresses, data from the *Building and Dwelling Register* can be linked to the residence of a person and the workplace of a business.