**Economic Commission for Europe**

**Conference of European Statisticians**

**Group of Experts on Business Registers**

**Twelfth session**

**Paris, 14-15 September 2011**

**Session 1: Linking statistical business registers across agencies, statistical domains and among countries**

# Linking business register data with other administrative data. An application to entrepreneurship indicators in Denmark

**Note by Dorte Høeg Koch (FORA, Denmark) and Mariarosa Lunati (OECD)**

**Summary**

The note presents an example of policy-relevant indicators developed by linkıng business register data with other administrative data. The example relates the experience of Denmark, where the matching of enterprise and personal data has allowed the development of statistics on the profile of entrepreneurs, i.e. their characteristics, competencies and qualifications. From a policy perspective, the indicators generated are very informative as they can support analysis and policy design in important areas such as women entrepreneurship.

## 1. Introduction

In the past decade there has been a rising policy focus on new enterprises, their contribution to job creation and innovation, and their possible impacts on the economy and society. The positive outcomes of the entrepreneurial activity can be many, encompassing in addition to innovation and employment creation important policy areas such as regional development, social inclusion and poverty reduction. For a thorough understanding of entrepreneurship and its impacts, and in order to establish a favourable environment for it, relevant and accurate statistics should be available.

In response to this need, the OECD-Eurostat Entrepreneurship Indicator Programme was launched in 2007, to develop policy-relevant and internationally comparable indicators of entrepreneurship and its determinants. The original approach of the programme consists in producing entrepreneurship indicators based on business register data, with the support of a Manual that provides internationally-harmonised definitions and methodology for the data treatment. The output of the programme is a new international collection of indicators of business demography (birth, survival, growth and death of enterprises). Constructed entirely from business register data, the indicators provide original information on the entrepreneurial performance of countries; this has proved very valuable for policy analysis. There are questions, however, that cannot be answered by exploiting only business register data, such as the influence that specific characteristics of the entrepreneur have on the success of a business. Based on the Danish case, the object of this note is to explain how much-needed data on entrepreneurs can be generated by linking the business registers with other administrative data. But beyond the example of entrepreneurship data, the note aims to point to the potential of data linking in general.

The following sections explain how Denmark has used the administrative registers for producing statistics. Due to three unique identifiers (of businesses, individuals and buildings) and a visionary law, data can be linked in Denmark fairly easily. This has made it possible to create a longitudinal database on entrepreneurship, which consists of data on both enterprises and persons. The note describes how the data linking is realised in Denmark, and highlights the main challenges of doing the linking. It emphasises how the formation of a very rich database, Entrepreneurship Database, has served the needs of evidence-based policy making in the area of entrepreneurship. The final section concludes with remarks on the relevance of the linking exercise and the key problems experienced in Denmark with respect to it.

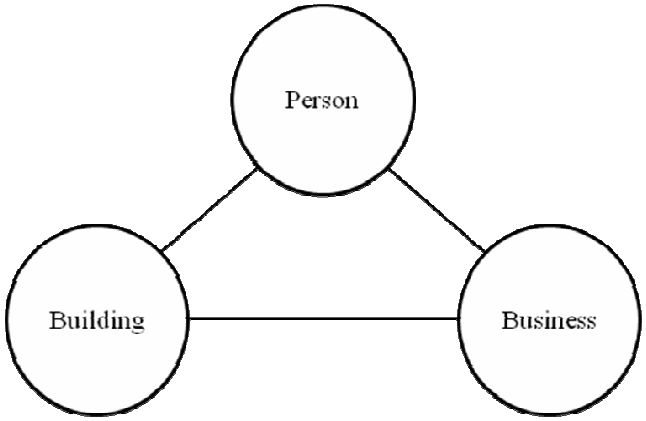
## 2. Linking data in Denmark

Denmark introduced a personal identification number – the so called person number – in 1968. But it was not until 1981 that the registers were sufficiently comprehensive and well established to be used for a proper register-based population census. This was due, among other things, to a less developed IT-system at that time.

In 1999 a new central business register number (business number) was introduced. The new business number replaced two old identification numbers. Recently the Building and Dwelling Register has been improved, in particular as concerns the address code that constitutes the identification number for buildings and houses. This code is important for the geographical break down of data.

These three different unique identification numbers are used in all administrative registers. This - combined with a visionary statistical law from 1966 - has made it possible for Statistics Denmark to produce the vast majority of its official statistics based on registers. In addition the identification numbers also allow for a multitude of data combinations used for inter-disciplinary issues.

**Figure 1. The three basic identifiers for building official statistics in Denmark**




The construction of both the person number and the business number is such that their validity can be controlled by a mathematical calculation (modulus 11-check). The person number incorporates information about the person – namely the birthday (dd mm yyyy) and the gender. It is important that the identification numbers are able to identify the units uniquely.

The strength of the system is that the identification numbers permit to link data from different registers easily – both within a specific year but also longitudinally. This makes it possible to gather many different kinds of information and also to create new variables that accumulate and/or combine information. For example, Statistics Denmark has created indicators to measure for the overall working experience of employees (i.e. how many years have they worked in their lifetime) and in a given enterprise. It is also possible to link data about a person prior to some event (e.g. starting-up an enterprise) in order to analyse his/her initial status. The linking over time has allowed identifying the continuing establishments via a definition built on the share of employees that are the same in two following years.

The identification numbers are used in business surveys as well; importantly both public and private research institutes use the identifiers when conducting surveys. Therefore information gathered through surveys can be linked to statistical databases that are used for producing the official statistics. The construction thus makes it possible to produce “inter-disciplinary” statistics of all kinds.

It is important to emphasise that the linking of data about persons are straightforward. However, the linking of data about businesses can be more problematic. The main reason is that businesses experience continuous change. For example, enterprises may create separate units for specific functions (i.e. the sales department) and these units may appear in the register as new and independent. Enterprises also buy and sell new establishments. They change their production. They move jobs to other countries. And so on. These many changes make it difficult to identify enterprises uniquely – especially over time. An important part in producing statistics on administrative registers is therefore to have events, and date for the events, reported as well.

The administrative registers have a very high data quality and the statistics from these sources are considered very reliable. As the registers are including all the relevant units there are no problems with non-response, sample size etc.

A challenge about using administrative registers is to keep definitions unchanged over time. This applies especially to new ways for enterprises to report information. Statistics Denmark is currently working together with the tax authorities following the introduction of a new VAT form that generates some inconsistent data about exports. This problem could result in a data break. This type of difficulty arises sometimes when additional data are requested for statistical purposes only to authorities that gather administrative data; the additional data provided are not always the object of a careful control.

In addition, it should also be considered that registers only provide "hard data", while for analytical purposes opinions, motivations etc. may be needed. With business register data it is not possible, for instance, to know what motivates entrepreneurs to start-up their business. These kinds of data have to be gathered through a survey.

Finally, the huge amount of gathered and linkable data can make parts of the society unsecure and afraid of data misuse. For that reason, in Denmark there is a big focus on data protection and confidentiality – as in many other countries. A special Danish rule establishes that data under the responsibility of Statistics Denmark can never be used for administrative prosecuting a single enterprise or person. Therefore, when Statistics Denmark recently discovered that there were companies that by law should have a board but did not have one, the finding could not be used to seek and find these companies. The responsible agencies must themselves solve the problem.

## 3. Exploiting the linked data

FORA – a division for research and analysis at the Danish Enterprise and Construction Authority has access to microdata; the main interest is on business data, and also data on entrepreneurs and employees. FORA makes an intensive use of microdata, as the Authority's strongly promotes evidence-based policy making.

In the recent years growth and job creation in new enterprises have been at centre of much policy attention in Denmark as well as many developed and developing countries. Danish policy makers are particularly interested in initiatives promoting the creation of growth enterprises. In order to design the most effective programs towards new growing enterprises, having access to data on new enterprises can be very useful.

Statistics Denmark has produced a longitudinal database on new enterprises since the 1990s. The basic foundation for this is the business register. The statistics are based solely on the use of administrative registers. The longitudinal dimension of the database allows to follow the enterprises over time in order to analyse survival and growth.

The population in this statistics is composed of new enterprises in "marked-oriented" industries, established under four legal forms (e.g. sole proprietorship, partnership and limited partnership, private limited company and public limited company). There are around 18-24.000 new enterprises every year.

In the case of new enterprises created under the legal forms of sole proprietorship, partnership and limited partnership, the business register contains a person number for the entrepreneur(s). For the private limited companies and public limited companies, the information about the establisher(s) is instead linked with a register for companies owned by the Danish Commerce and Company Agency. For a few of these new companies it is not possible to identify one person as the entrepreneur. This is mainly due to the way the paper work related to establishing a new company is prepared; in Denmark, lawyers and accountants may be in charge of it.

With the person number of the entrepreneur the statistics can be broken down by person-specific indicators. In addition, it is possible to link the employees (with person numbers) to the enterprise and link the indicators to them as well. The construction therefore allows for many cross-tabulations, as illustrated in Table 1.

**Table 1 Content of the Entrepreneurship Database of Statistics Denmark**

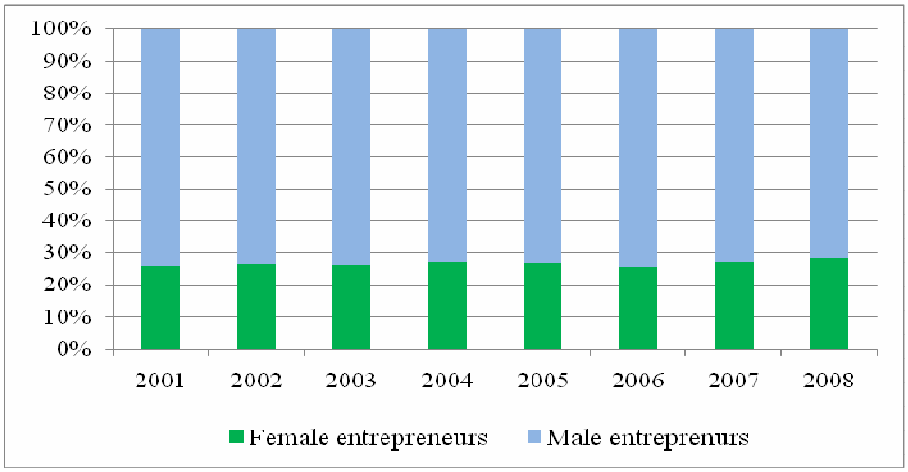
| New enterprises, surviving enterprises, gazelles | | Entrepreneurs, employees, (board members) |
|---|---|---|
| • industry (NACE+2 Danish digits)<br>• size class (employees)<br>• legal form<br>• geography<br>• turnover<br>• export<br>• number of employees<br>• (destination countries of export)<br>• (board) | x | • gender<br>• age<br>• nationality<br>• education<br>• Marital/family status<br>• number of children (pre-school age)<br>• age of children<br>• parents occupation<br>• total work experience<br>• industry experience<br>• leadership experience<br>• job status the year prior start-up<br>• gross income the year prior start-up<br>• (international experience)<br>• etc. |

For the surviving enterprises and gazelles it is assumed that the entrepreneur is the same after five years as when the enterprise started. This makes it possible to analyse if the initial entrepreneur's characteristics, in terms for instance of qualifications and experience, are important for to determine a business' success.

In brackets are also mentioned board members and a few indicators. At the time being these are on development. The board members have been identified for all companies with less than 100 employees; it is planned to analyse the importance of having a competent board for the performance of smaller businesses. Denmark is also investigating if employees with international experience have an impact on new businesses' ability to export.

It is also possible to create statistics on new enterprises by gender (Figure 2). The share of female entrepreneurs is almost unchanged in Denmark since 2001; it seems therefore that there is a huge potential to get more women to start business and think in terms of growth. Today, less than a fourth of the new enterprises in Denmark are started by a woman. Though, women have the skills and knowledge to start successful businesses, as more women today carry out a longer education than men.

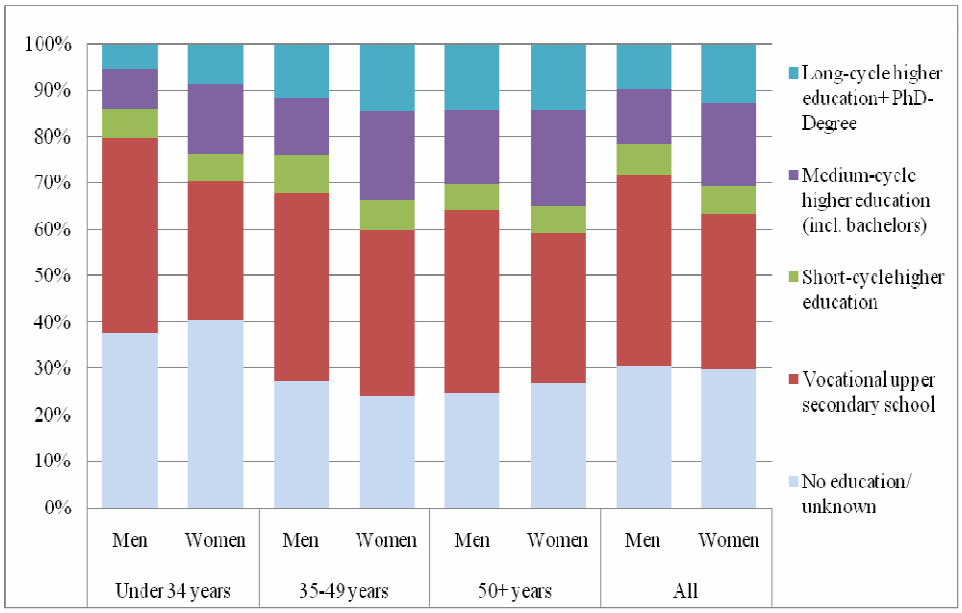
**Figure 2. Entrepreneurs by gender, 2001-2008**




*Source*: Own calculations based on data from Statistics Denmark.

Even though in 2008 the share of female entrepreneurs with higher education for all age groups of entrepreneurs had increased compared to the past (Figure 2), it is still considered important in Denmark to ensure that in the future more young well-educated women will participate to the creation of new businesses.

**Figure 3. Entrepreneurs by gender, age and education, 2008**




*Source*: Own calculations based on data from Statistics Denmark

To that end, in recent years the Danish Enterprise and Construction Authority has introduced an action plan to encourage women entrepreneurship, which was developed in collaboration with women entrepreneurs. The design of the program benefited considerably from the analysis of the linked data; the findings concerning the characteristics of women entrepreneurs and the business they owned were critical to the setting of the policy initiative. Although the purpose of this note it is not to comment on the specific trends in women entrepreneurship in Denmark, some of the main findings of the analysis of the gender data are presented below, as an illustration of the richness of the linked data. The main findings included:

- There is no difference between the share of male and female entrepreneurs that become gazelles.
- Women establish their businesses when they are in their most fertile age. Therefore they have to be able to cope with maternity and small children on the one hand and the workload that comes with a new enterprise. There appear to be big differences in business survival rates between industries when it comes to female entrepreneurs with small children. It is more difficult in retail trade than in knowledge-intensive types of activities.
- Regarding the distribution by industrial activities, many women start up in retail trade, which has no or low entry requirements and it is therefore a very competitive sector – as most of the other typical female industries are.
- 50 percent of Danish women are employed in the public sector. Econometric analysis showed that this is the main reason for the lower degree of female than male entrepreneurs in Denmark. It can also be seen than most of the female entrepreneurs actually come from the private sector as the public sector is not where you get the competencies for starting up a new business.
- More women than men start a new business in an industry where they have no prior experience. This is a problem, because survival and growth are generally correlated with previous industry experience.

**Table 2 Entrepreneurs 2008 by gender and industry experience**

| | Percentage | | |
|---|---|---|---|
| | Men | Women | Total |
| Industry experience | 37,8 | 33,4 | 36,6 |
| No industry experience | 51,2 | 50,2 | 50,9 |
| Unknown industry | 11 | 16,3 | 12,5 |
| Total | 100 | 100 | 100 |

*Note*: Data cover only new enterprises where data can be linked to persons.
*Source*: Own calculations based on data from Statistics Denmark

Finally one result appeared of special interest: the only significant indicator to explain why women choose to be an entrepreneur was that their mother had been self-employed. This result was used to start a campaign of role models and to make successful women entrepreneurs as visible as possible. The campaign was launched in connection to the Global Entrepreneurship Week (a worldwide event to promote the knowledge of and interest in entrepreneurship) and has been repeated every year since.

## 4. Conclusions

This note highlighted the potential, for analysis and policy design, of data obtained by linking business registers with other administrative data, as the linked data are of high quality and very flexible. The note focused on a specific category of linked data, developed in Denmark to analyse entrepreneurial activities. The Danish Entrepreneurship Database represents a powerful tool for policy makers in Denmark: the rich and detailed data contained in the database make it possible to analyse many characteristics of entrepreneurs in general and of special target groups in particular, i.e. female entrepreneurs.

In addition, the intensive use of registers in producing statistics is also a less costly method than using surveys. A review of Statistics Denmark in 2007 conducted by three internal experts on behalf of the EU pointed out that “the efficiency of Statistics Denmark is very high – the main reasons are broad sharing of administrative data sources and use of modern statistical methods”. In addition, the expert

highlighted that "the administrative burden on businesses emanating from statistics is very low in Denmark. In 2005, it was just 0.5 percent of the total administrative burden. The extensive use of data from administrative sources is the reason for this low percentage."

For Denmark there are, though, still two main challenges: the lack of timeliness and the possibility of benchmarking with other countries. In relation to timeliness, and still in the area of entrepreneurship data, the most recent statistics on entrepreneurs refers to 2008. This is a non negligible limitation, in particular in light of the global financial crisis. More recent data would be needed to understand how the crisis has affected the entrepreneurs and whether changes in the support programs are necessary. FORA is therefore looking for other data sources that can be used as preliminary information.

While much could be learned from cross-country comparisons, the scope for analysis of entrepreneurs and their business at the international level is still very limited. Only a few countries produce statistics comparable to those generated by Denmark through linking the business register with other administrative data. With more efforts in the direction of data linking, a number of topics that are of high policy interest for growth and job creation could be usefully investigated across countries, such as the survival and growth of enterprises created by women, the performance of entrepreneurs with foreign background, the evolution in the number of young entrepreneurs and entrepreneurs with higher education.

And the political interest does not stop at the entrepreneurs. Comparable business statistics that are linked with data on persons can be useful in many other contexts.

**References**


Borchsenius, Lars (2005): New developments in the Danish system for access to micro data, Joint UNECE/Eurostat work session on statistical data confidentiality.

Bøegh-Nielsen, Peter (2001): Statistics on start-ups and survival of women entrepreneurs: The Danish Experience, in *Women Entrepreneurs in SMEs, Realising the Benefits of Globalisation and the Knowledge-based Economy*, OECD

Erhvervs- og Byggestyrelsen (2008): Kvinder kan – få succes med egen virksomhed, www.ebst.dk

Erhvervs- og Byggestyrelsen (2009): Handlingsplan for Kvindelige Iværksættere, www.ebst.dk

Eurostat-OECD (2007): Manual on Business Demography Statistics.

Junge, Martin (2009): Vækstiværksættere. CEBR, Unpublished paper

Matějček, Jan, Inger Öhman (chair), Paloma Seoane Spiegelberg (2007): Peer review on the implementation of the European Statistics Code of Practice. www.dst.dk

OECD (2011): Entrepreneurship at a Glance.

Spieker, Finn (1999): Access to confidential data in an integrated statistical system, Joint ECE/Eurostat Work Session on Statistical Data Confidentiality.

Statistics Denmark (1994): Personstatistik i Danmark

Statistics Denmark (2006): Statistical Code, www.dst.dk

Statistics Denmark (2010): Numbers on time, www.dst.dk

Sundgren, Bo and Svein Nordbotten (2003): Review of Statistics Denmark, Norway 30. April 2003, www.dst.dk