

Working Paper No. 29 (Summary)
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

Joint ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (ii): New data release techniques

**FROM ON-SITE TO REMOTE DATA ACCESS –
THE REVOLUTION OF THE DANISH SYSTEM FOR ACCESS TO MICRODATA**

Contributed paper

Submitted by Statistics Denmark¹

¹ Prepared by Otto Andersen (oan@dst.dk).

ABSTRACT

Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality Topic (ii): New data release techniques

From On-Site to Remote Data Access, the Revolution of the Danish System for Access to Microdata

Statistics Denmark has altered its scheme for giving researchers access to de-identified micro data from on-site to remote access through the Internet. This is part of the general vision that Denmark should work hard to be one of the Worlds leading countries within register-based research. Through the new scheme Danish researchers have experienced a breakthrough in the methods of access to micro data.

From surveys to register based statistics

In the mid-1980s, Statistics Denmark experienced an emerging interest among various research environments and ministerial analysis divisions in applying micro data (individual data) for research and analysis purposes. At that time the Danish statistical system changed completely to be based mostly on data from administrative registers instead of surveys and the development in computer technology made it technically possible to process large amounts of data according to advanced statistical models, such as multivariate models.

The contents of the Danish statistical registers cover many fields of research such as labour market research, sociology, epidemiology and business economics. The strength of the system is that the identification keys (Person Number, address, central business register number and property title number) render it possible to correlate the aggregated data both within a specific year and longitudinally across several years. The data potential for research is unique.

Confidentiality

Statistics Denmark is not allowed to hand out micro data because of the rules of confidentiality lay down by the Management and Board of Statistics Denmark. However, it was evident already at in the mid-1980s that not only were the registers of enormous importance to Statistics Denmark, but their research potential was so great that it would be very valuable to actually utilise them for research purposes. Therefore, Statistics Denmark had to find a solution to the problem of access, which complied with the existing legislation on registers while taking into account Statistics Denmark's own confidentiality principles.

The on-site arrangement

The solution was to establish an on-site arrangement allowing researchers to get access to de-identified micro data from work stations in Statistics Denmark. The individual researcher signs a contract with Statistics Denmark stipulating first of all the confidentiality rules. Even if the researcher only have access to de-identified data (the Person Number is substituted by a serial number) the data are normally so detailed that they are not anonymous. Therefore data are considered by Statistics Denmark as confidential and no information on which could lead to the identification of persons or firms must be published. The researcher can, however, use the micro data to produce statistical tables etc.

The Danish research environment has often complained about the difficulties due to the geographical distance to Statistics Denmark. In order to give easier access to data for researchers resident in the western part of Denmark a small branch of Statistics Denmark was established in Århus, Jutland offering an on-site arrangement as in Copenhagen.

Furthermore special computers have been installed in Statistics Denmark only for the use of researchers working on-site. The platform was chosen to be UNIX on SUN-computers.

Despite the difficulties the on-site arrangement has been used very frequent. In 2001 more than 150 researchers had a contract with Statistics Denmark.

The remote access arrangement

During the autumn of 2000 Statistics Denmark initiated a pilot study in order to develop a remote access system giving the same security as the well tested on-site system. It was found that such a system could be established and was started up 1 March, 2001.

Access to data is only granted to researchers within an (by Statistics Denmark) authorized research institution. The amount of data is restricted according to a “need to know” principle meaning that the researcher only has access to the data needed for the specific project. At a start very sensible data especially enterprise data were excluded from the scheme.

Micro data are stored on the above mentioned SUN/UNIX servers in Statistics Denmark. The technical solution is based on Citrix servers and the use of the Internet. Communications via the Internet is encrypted by means of a so-called RSA SecurID card, a component that secures Internet communications against unauthorised access. In practice the researcher rents a password key (a token) from Statistics Denmark. The token ensures that only the authorised person obtains access to the computer system. The functionality at external access is largely the same as on the PCs under the on-site arrangement. Printing and data transfer options are not available.

Printouts are sent to the researcher by e-mail, logged at Statistics Denmark and checked for confidentiality problems by Statistics Denmark.

The latest development

The remote access arrangement has been evaluated after 1.5 years and Statistics Denmark is convinced that the security around the scheme is adequate and at the same level as in the on-site arrangement. Therefore Statistics Denmark will grant remote access to all accepted (“need to know”) project data to authorized Institutions.