Working Paper
ENGLISH ONLY

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)

CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Helsinki, Finland, 5 to 7 October 2015)

Topic (iv): Access to Statistical Data for Scientific Purposes

New Nordic model for researchers joint access to data from the Nordic Statistical Institution

Ivan Thaulow* and Charlotte Nielsen*

* Division of Research Services, Statistics Denmark, Sejrøgade 11, DK-2100 Copenhagen, ITH@dst.dk.

Parts of this article is excerpted from the report "Feasibility study regarding research access to Nordic Microdata", Statistics Denmark, 2014.

Abstract: The Nordic National Statistical Institutions (NSI's) have agreed that researchers access to social micro data from different countries may be given through already existing remote access systems in the Nordic NSI's - at the moment in Denmark, Finland and Sweden. Thus micro data from different countries may be collected in one NSI and from here accessed remotely accessed by researchers from any of the Nordic countries. The present paper will describe the Nordic model some selected data confidentiality issues.

1 Introduction

In the Nordic countries there has been a long tradition of collecting data for statistical purposes from administrative systems. This tradition has led to data covering the entire national population by means of long-term data series of high quality, which are stored and used as basis for compiling statistics. All Nordic Statistical Institutions (NSI's) give national researchers access to de-identified microdata and register-based research is prominent in most Nordic countries. However, cross Nordic register research is still rather rarely carried out.

Within these premises the Nordic research community has expressed a need for improved possibilities of joint access to Nordic microdata, which could ease the efforts necessary to carry out analyses involving register data from more than one country. The mechanism should aim to make access easier, as regards administrative procedures, communication and information about procedures, as well as technical and economic conditions.

In 2014 a Task Force from Nordic National Statistical Institutions (NSI's) carried out a feasibility study regarding joint access to Nordic microdata for researchers. The most significant output from the work done by the Task Force was a Nordic model for researcher's joint access to data from the Nordic Statistical Institution. The Nordic Chief Statistician was presented for the model and decided to test and evaluate the model for joint access to research microdata in the next years. During that period the consequences of the model will be monitored closely and changes in the model will be made along the way if needed.

2 National practices for researchers' access to microdata in the Nordic countries

As part of its work the Task Force made a review of the legal frameworks for researchers' access to microdata as well as a review of national practices for researchers' access to microdata in the Nordic countries.

2.1 Legal Framework

The legislation concerning confidentiality and protection of an individual's integrity is the foundation for how access to data for research is processed by the Nordic National Statistical Institutions. The basic European legislation currently covering data protection within the EU is the Data Protection Directive 95/46/EC. In general, the laws and regulations of relevance for researchers' access to microdata in the Nordic countries are the implementation of this EU Directive. However, national differences in how the EU Directive is implemented in National laws and practices exist. In addition, other national legislation, e.g. Statistical Acts and/or acts about public information and secrecy have to be complied with.

However, the review of the legal framework concluded that the Nordic countries have a similar legal basis to give researchers access to microdata and that legislation in all Nordic countries does make it possible to give researchers access to microdata. It also allows the transfer of de-identified data from one NSI to another under certain conditions.

2.2 Type of access to microdata and administration of access

Remote access to de-identified microdata is currently offered by four Nordic NSI's (Denmark, Finland, Sweden and Greenland). In all remote access systems the de-identified microdata are kept safely behind firewalls, and data can only be accessed through safe, encrypted lines (Table 1).

However, some countries still use on-site access and even hand out data to researchers but the latter ways of giving data access will probably be phased out in the years to come, e.g. Norway is making plans to implement a remote access system (the RAIRD system).

Table 1: Researchers' access to microdata – type of access to microdata and administration of access

	Denmark	Finland	Sweden	Norway	Greenland	Iceland
Provide microdata for research?	Yes	Yes	Yes	Yes	Yes	Yes
Type of access to microdata						
Remote access	Yes	Yes	Yes	(No)	Yes	No
On-site access	No	Yes	No	No	No	(No)
Data handed over to researcher	No	Yes	(Yes)	Yes	No	Yes
Administration on access						
Authorization needed	Yes	Yes	Yes	Yes	Yes	Yes
Public companies	Yes	Yes	Yes	Yes	Yes	Yes
Private companies	Yes	Yes	Yes	Yes	Yes	Yes

2.3 Data security and data confidentiality

In all Nordic NSI's an important aspect of data security, which is also covered by the law, is to provide researchers access to only the data they need in order to complete the study in question – the "need-to-know"- principle. The **"need-to-know" principle applies both to the population, registers and variables.** As shown in table 2 (next page) all Nordic countries follow this practice.

The Nordic countries have fairly similar authority structures and high levels of security in the way researchers are given access to microdata. Thus, in all Nordic countries access for researchers is only given to **de-identified data** and always after conducting **a case-by-case evaluation of which data are needed** in the research projects (the so-called principle of "need to know") as well as the risks of damaging and harming the research objects (Table 2).

Furthermore, all researchers and/or environments have to sign a **security agreement** before access to microdata can be granted, making sure that the researchers/environments are fully aware of the possibilities and limitations in using microdata for research and aware of the consequences if the agreement is broken.

All countries with remote access apply **output control** on analyses performed by researchers. However, some different procedures for output control exist among the NSI's. In Finland all output is controlled before it is handed over to the researcher, Denmark and Sweden have random output control after handing over the output to the researcher.

If a security breach is discovered **all countries are able of imposing sanctions -** either administrative or legal sanctions – against an offender depending on the type and severity of the security breach.

Table 2: Measure for data security and protection of personal information

	Denmark	Finland	Sweden	Norway	Greenland	Iceland
Before access						
Restrictions	"Need to know"	"Need to know"	"Need to know"	"Need to know"	"Need to know"	"Need to know"
Security agreement signed by	Researcher and research environment	Researcher and research environment	Researcher and research environment	Research environment	Researcher and research environment	Researcher
De-identification of data	Yes	Yes	Yes	Yes	Yes	Yes
Disclosure control	No	For data that are handed out	Yes	Yes	No	Yes
After access						
Output control	Random sample	Complete	Random sample or complete if logins from outside Sweden	Not relevant	On suspicion	Not relevant
Sanctions if security agreements are broken	Possibilities for administrative and legal sanctions	Possibilities for administrative and legal sanctions	Possibilities for administrative and legal sanctions	Environment can lose eligibility to access microdata and legal sanctions	Possibilities for administrative and legal sanctions	Possibilities for administrative and legal sanctions

2.4 Data available

The review performed by the Nordic Task Force also showed some differences between the Nordic systems of research services. A conspicuous national difference is seen with regard to the ownership of certain social data differs between the countries. Comparable data might be owned by the NSI in one country and by separate institutions in other countries (e.g. data on sick leave) with the implication that not all NSI's are able to permit research access to these data. Thus, other institutions have to be requested to give permission to access before these data can be included in a research project. Similarly, national differences are seen regarding what is considered as sensitive data that calls for approvals from ethical committees.

Some differences can also be seen in how the data access systems are organised in practice. For instance, some NSI's have centralised units of Research Services to handle both the requests of researchers and the extraction of the relevant data while others have a more decentralised way of handling these tasks.

3 Outcome – a model for cooperation between Nordic NSI's

Based on the similarities in the Nordic research arrangements, the Nordic NSI agreed in September 2014 on the described joint model of cooperation for access to Nordic microdata for research. It will be tested and further developed within the next couple of years.

In the model of cooperation, some basic steps of the progress of a Nordic research project were agreed on. Due to slightly different administrative rules in the Nordic countries - the order of the steps may differ.

Step no.	Description
(1)	Approval/Authorization of researcher and/or research environment by each NSI
(2)	A project application for access to microdata for a specific research project is received in each NSI
(3)	Security agreements between researchers and data hosting NSI
(4)	Approvals from relevant national authorities
(5)	Evaluations and approvals of projects from each involved NSI's
(6)	National costs are calculated and contracts between principal researcher and NSI's are signed
(7)	Data exchange agreements are signed between the relevant NSI's
(8)	Extraction of data by each NSI
(9)	Disclosure control by each NSI
(10)	Data transfer from Nordic NSI's to the data hosting NSI
(11)	Access to data through the remote access system of the datahosting NSI
(12)	Output Control by the datahosting NSI
(13)	Common handling of security breaches (initiated by the datahosting NSI)
(14)	Closing the project at the datahosting NSI

The first 10 steps are carried out by each of the "local" NSI's, while the last four steps are implemented primarily by the datahosting NSI.

The most important achievement of the model is that it has been agreed that de-identified microdata from one NSI, may now be transferred to another Nordic NSI in order to include and analyze data from multiple countries simultaneously and thus test scientific questions on a more solid ground. Thus, the model makes it possible to include data for up to 25 million people in the six Nordic countries, if needed.

Data available for the new Nordic model includes all "Social Data" in the NSI's. That is all data in the NSI regarding their population, living conditions, labour and unemployment, earnings and income, education etc. Health data and data about enterprises are not included in the model yet, so if researchers need these kinds of data for their projects special agreements have to be made about this.

Another important advantage of the model is that certain common administrative processes and security rules are now described and agreed on. However, it has to be pointed out that in the model of cooperation, it is, crucial that the national sovereignty to decide if researchers can get access to their national microdata and for which purpose – is kept unchanged. So, all decision-making on national data still remains within each NSI. When access has been approved according to national laws and regulations in each country, all relevant data are extracted, de-identified and billed by each NSI separately.

It is a prerequisite for providing access to cross Nordic microdata through one Nordic NSI that data security is given the highest priority. Therefore, a common Nordic security agreement is part of the model. It has to be signed by all participating researchers as well as each relevant NSI.

In the Nordic model of cooperation access to Nordic microdata can only be given through the existing remote access systems. Thus it has been agreed to use the three existing remote access systems in Finland, Sweden and Denmark to give Nordic researchers access to data from all 6 Nordic countries, since remote access seems to be the best way to combine a high level of data security with high usability. In brief - **data never leaves the Nordic the safe domains of The Nordic NSI-system**, it can only be accessed through encrypted, safe connections, it is safeguarded by our firewalls, the researchers' output is logged by the data-hosting NSI and the content of all output is controlled.

4 Conclusion and perspective

The Nordic model will make it easier for the researcher society to work with microdata from multiple Nordic countries in the future. The researchers are now – through a remote access in just one Nordic NSI – able to work on-line with relevant data from all Nordic countries.

The Nordic model is based on the present national research arrangements in the Nordic countries with some new standard procedures in handling a research requests for data from several Nordic countries. An important element of the model is that it ensures a high level of data security.

In the agreed model, data can be accessed from an NSI's with a remote access system and a common security agreement is drafted, an application form, and agreement for exchange of data between the NSI. Furthermore, common guidelines in handling security breaches have been agreed upon by the NSI's. The model also ensures a structured cooperation and communication between the Nordic NSI's when handling a cross Nordic research project. In this way one can speak of a somewhat centralized model.

Though the purpose of the model is to make it easier for researchers to make cross-country comparisons, the model is rather winded. This is due to the fact that the model is still fairly de-centralized, since all important decisions about national data are still made within each NSI. This means that 1) aall approvals for access (users as well as projects) to data are kept at national level, and 2) all selection and preparation of national data will be carried out by the each NSI. This include: guidance, data extraction, de-identification and transfer of data to the data-hosting country. Finally all prices attached to national data are calculated nationally.

Since the Nordic model greatly respects the national needs to keep control over national data, it might be of interest to other countries or international organizations that want to promote cross-country comparisons.

Before trying to implement elements from the Nordic model in other regions of Europe, one has to take two things into consideration:

The first is, that the success of the Nordic model is based on a strong foundation of ccultural and linguistic similarities in the Nordic countries, we all have well established welfare states and democracies, a long historical tradition for Nordic corporation, extensive similarities in the political/administrative system surrounding data use and data protection, rather similar legal basis to give researchers access to data and corresponding high levels of security in the way researchers are given access to micro data. If we have not had all this positive background in common, the task of finding a common model would surely have been much more difficult.

Finally, it is important to stress that although written common rules and agreements must be an important part of a common model, such rules and agreements cannot be made exhaustive. Reality is always far more complex that even the most forward-thinking lawyer can anticipate. Therefore, trust between the cooperating organizations is of the essence. We have to thrust that each of us does as we have agreed on, and we also have to trust, that we will find common acceptable solutions, on the problems that we cannot yet foreseen. So, trust has to be the starting point of joint model, and trust has to develop as we move along.