Sharing registry data for health research in the Nordic countries – a proposal for increased collaboration

Report from the Nordic Task Force for Access to national data repositories

1. Magnus Stenbeck, Ulf Jonsson, Sweden, Maria Heimisdottir, Iceland, Mads Melbye, Denmark, Marjut Salokannel, Finland, Camilla Stoltenberg, Norway.
Contents

Contents ........................................................................................................................................ 2

Executive Summary - High priority recommendation ............................................................... 3

This report .................................................................................................................................... 4

The mandate .................................................................................................................................. 4

The Task Force ............................................................................................................................ 5

The Task Force’s recommendations ............................................................................................... 5

NCHD Mandate ............................................................................................................................ 6

Allocated resources and initial operation ..................................................................................... 7

Further recommendations ............................................................................................................. 8

Legal changes and political agreements on data sharing ............................................................ 10

E-science solutions - Technical development of a distributed data infrastructure .................... 15

Appendices

Appendix 1. The Nordic Conference on Registry Epidemiology

Appendix 2. Examples of Nordic Collaboration

Appendix 3. Open letter to decision makers

Appendix 4. Video recordings from the Nordic Conference (8 files)
This report presents the proposal of NordForsk’s Expert Task Force on Access to National Data Repositories. It outlines the main components of a sustainable Nordic infrastructure for sharing national register data for health related research.

**Executive Summary - High priority recommendation**

The Nordic countries have unique health registries and databases. In order to facilitate research on these resources, the Task Force proposes the following action:

1. A Nordic Center of Excellence (NCoE), i.e. the Nordic Center for Data Collaboration within the Health Sciences (NCDH), should be initiated by NordForsk in the spring of 2012.
2. The NCDH should be a network organization with a single physical node and an associated publicly available portal in each Nordic country (i.e. NCDH.DK, NCDH.FI, NCDH.IS, NCDH.NO, and NCDH.SE). The initial set up of national nodes should be completed by the end of 2012, and the operation of the national NCDHs should commence no later than January 1, 2013.
3. As from 2013, the national NCDH should act as a contact point for research that
   a. needs to get access to register data from neighboring countries, and/or
   b. needs to set up collaborations across the Nordic borders
4. The NCDH network should receive funding for its basic operation from NordForsk. The initial funding period should cover at least 3 years and consist of 1 FTE plus administrative resources per country. Additional funding and/or support may be added from national or other sources.
This report

The details of the above proposal, as well as the background work, are described below. The report commences with a description of the mandate given to the Task Force and its organization and work, after which the task force’s considerations and conclusions are described in detail. The report concludes with a discussion on legal issues and technical solutions.

In order to receive input to this report, the task force arranged a workshop on how to facilitate Nordic collaboration on the registry-based research within health and welfare. The workshop was held at the Nordic Conference on Registry Epidemiology in Iceland on June 15-16 2011. Six speakers with extensive experience of Nordic collaborative research within the health sciences were invited. A description of the workshop is attached to this report (Appendix 1), along with video recordings of the speeches (Appendix 4). In addition, illustrative examples of Nordic collaboration projects are attached (Appendix 2), as well as an "Open letter to decision makers" from Nordic researchers concerning the upcoming revision of the European Data Protection Directive (Appendix 3).

The mandate

In 2010 a Nordic Expert Task Force was appointed by NordForsk. It was given the mandate to look at possible improvements of researcher access to national data repositories. The Task Force was commissioned to:

1. Recommend concrete actions with an identified time plan that can be undertaken as a next step.
2. Describe which Nordic level functions and services would be beneficial in order to improve Nordic collaboration in registry based research.
3. Propose improvements in terms of information about data resources in the Nordic countries and how they can be accessed.
4. Present an inventory of the legal obstacles and possibilities in the Nordic countries and a comparison between them. It should propose a common strategy of legal reform for facilitating research on sensitive personal data that maintains or strengthens the personal and commercial integrity protection for the research subjects.
5. Describe further actions needed on the Nordic and national levels, such as investigating the possible technical solutions involved in providing access to a common system of shared government and researcher owned data for research purposes.

6. Evaluate the need for a demonstration project using a limited number of data sources within an already well defined area of research.

7. Describe the expected benefits of a joint Nordic collaboration from a wider international perspective.

The Task Force

The Task Force has consisted of the following national members:

Maria Hemisdottir, Iceland
Mads Melbye, Denmark
Marjut Salokannel, Finland
Magnus Stenbeck, Sweden, chair
Camilla Stoltenberg, Norway

In addition, the Task Force had assistance from Katherine Svensson as secretary (2010) and Ulf Jonsson (2011) as a report author and organizer of the Icelandic workshop. Several national experts have contributed during the project. Within itself, the Task Force possesses expertise within epidemiological and social research, law, and civil service related to public health monitoring and the administration of national registers.

In 2010 the task force had three physical meetings, in Stockholm and Oslo. In addition, several telephone conferences were held in 2010-2011. Members of the Task Force participated in the organization of the Nordic conference on register based health research 2011, and met in Reykjavik to draw conclusions from the conference.

The Task Force’s recommendations

excellence) for the purpose of collaborative research and data sharing. In this report, the center is labeled the Nordic Center for Data sharing in Health research (NCDH).

2. Its overall purpose is to facilitate exchange of data and research collaboration in the Nordic countries.

3. Initially, the NCDH should focus on collaborating within the health sciences, but other areas may be included eventually, possibly by creating sister organizations within the social sciences, climate and environment, and other fields.

4. The NCDH should collaborate with the Nordic Initiative on Biobanking and Molecular Research (BBMRI Nordic).

5. The NCDH should be a network organization with an independent node in each country. The node should communicate with the research community via a national portal set up under a joint Nordic umbrella (i.e. NCDH.DK, NCDH.Fi, NCDH.IS, NCDH.NO, and NCDH.SE).

6. The national node should provide assistance with practical and legal matters in Nordic research collaborations, ranging from advice and consultations to taking care of the practical handling of data transfers (depending on the allocated resources).

7. The Task Force recommends that the national node is set up within an already existing academic unit, data coordination facility, or government office willing and able to take on the responsibility to build a platform for Nordic research collaboration. If set up as a unit within such an organization, measures should be taken to guarantee that there is no conflict of interest between the purposes of the host organization and the NCDH goals. The resources for the NCDH should be clearly separated from other resources of the host organization, and it should be governed by its own steering committee.

8. A Nordic advisory committee should be appointed. The steering committee should have one representative per country and be chaired by a NordForsk representative.

**NCHD Mandate**

The mandate of the national portal should be to:

- handle contacts, advise individual researchers, and distribute data to and from researchers in their own country

- advise researchers on international and national rules and procedures for data access in other Nordic countries
✔ aid researchers in obtaining permissions and getting access to data from other Nordic countries

✔ coordinate ethics approvals and other necessary legal documents for Nordic projects lead from their own country

✔ collaborate with neighboring countries national portals, including the exchange of information and data through the NCDH channels

✔ collaborate with the national portals in the neighboring countries, and exchange information and data through this channel

✔ keep working contacts with domestic data providers and researchers

✔ provide simple and easily accessible information on available national data sources and links to more comprehensive data descriptions and data owners

✔ provide easily accessible information on the legal conditions and practical procedures under which access to data from national repositories

✔ negotiate with national government agencies concerning access to data for Nordic collaborative project, and distribute the results back to the researcher via his/her national portal

✔ in general, facilitate contacts between researchers in neighboring countries

✔ set up a feedback mechanism for the documentation of Nordic collaborative projects

✔ in collaboration with other national portals investigate how to implement in practice the new European data protection legislation and the legal changes that will be the consequence of this, with the goal of proposing a common Nordic policy that is beneficial to register based research in general and to Nordic collaboration in particular.

Allocated resources and initial operation

Initially, the national portals for Nordic collaborations should employ one person per country, plus necessary administrative resources. The contact person should be responsible for setting up services, plan the national structure, run initial operations, implement working collaboration with other
centers, and explore/generate financial and other national support for the continued operation of the center.

The NCoE network should be financed by NordForsk for an initial period of three years after which the organizational structure should be evaluated and the details of a long term permanent structure should be considered.

Further recommendations

The Task Force has discussed a number of additional reforms and initiatives that would facilitate Nordic collaboration. It is expected that these recommendations will provide input to the NCDH and that concrete action plans with respect to the recommendations will be developed in this context.

Nordic level functions and services

✓ The NCDH should set up goals and negotiate principles for prompt data delivery for Nordic collaborative projects (e.g. 30-60 days), that are adopted by the data providers in each country

✓ National storage facilities for preservations of research databases and programs for generating such databases should be created, enabling the reuse of data developed in research projects

✓ The NCDH should aid in the process of creating joint datasets for analysis.

Information about data resources

✓ The NCDH should create a basic data catalog of Nordic data resources with associated simple data descriptions

✓ The NCDH should create a common syntax of common data definitions for a set of core variables in registry data in order to increase the comparability of data across sources

Legal obstacles and possibilities

✓ The NCDH should consider and propose changes of the legislation and/or legal practices in each country, focusing on ways to facilitate national as well as Nordic access to registry data for research. The Task Force for instance recommends evaluating the Danish facilitated ethical vetting procedures as a model for other countries.

✓ THE NCDH may propose models for the coordination of legislation and vetting procedures within and between countries
The NCDH may propose principles that make data sharing over computer based networks legal between countries in the Nordic region.

**Technical solutions**

- Modern technology provides new possibilities for data sharing under safe conditions. One may consider Nordic federated data rather than large merged databases. The use of cloud computing techniques for distributed data should be evaluated. Data protection requirements associated with modern database technology should be considered.

- The possibility of building legal informatics systems for swift procedures and equal legal treatment should be explored.

**Demonstration projects**

- The Task Force has not organized its own demonstration project. The possibility of a Nordic project on the evaluation of the “swine flu” (H1N1) vaccination in terms of coverage and its effects on pregnancy outcomes and narcolepsy was discussed. Some examples of Nordic collaboration are described in the report. Documentation of study protocols and data sharing methods in future collaborative projects should become a regular undertaking of the NCDH.

**Wider international perspective**

- The Task Force realizes that the unique possibility to collect comprehensive data on entire populations for use in research and social policy must be motivated in a wider international context. It depends on the ability of Nordic research to show the usefulness of this approach outside of the Nordic region. In many cases, the Nordic populations can provide information that is useful far beyond our region. Hence, the Nordic data should be accessible for researchers in other regions, and current projects that show potential in this area should be made known. The NCDH should engage in the dissemination of new knowledge in this area, and may initiate articles on related topic in journals and other media.
Legal changes and political agreements on data sharing

Introduction

The NordForsk mandate to the group states that

“the proposal should present an inventory of the legal obstacles and possibilities in the Nordic countries and a comparison between them. It should propose a common strategy of legal reform for facilitating research on sensitive personal data that maintains or strengthens the personal and commercial integrity protection for the research subjects”.

After the formulation of the mandate, and in further discussions with NordForsk representatives, the task stated above has been limited to deal with health related research on individuals. Hence, the protection of commercial ownership interests and their relation to principles of open access to research data and results are not dealt with in this report.

In addition, the legal situation of data protection for individuals is undergoing a major revision at the European level. It is expected that in the beginning 2012 there will be a proposal for a comprehensive EU regulation relating to data protection in the European Union, replacing the current Data Protection Directive 95/46/EC and other directives related to the protection of personal data. The purpose of the new legislation is a complete harmonization of data protection regulation in the European Union. An EU regulation is immediately applicable law in all Member States after having entered into force.

Hence, although we are referring to and describing the current situation in the Nordic countries, the group has decided to work most actively on the anticipated consequences of the new European legislation. It is anticipated that this legislation will make the legal situation similar, if not identical, across many European countries, and will bring the rules of the Nordic countries closer together.

A proposal for joint Nordic action on this is presented at the end of the section “Review of the European data Protection Directive” below.

Background

The basic principles for data protection are laid down in the UN human rights principles, Article 12, and in the Helsinki declaration, Article 22. The basic European legislation currently covering data protection within the EC is the Data Protection Directive 95/46/EC. Although it has a longer historical background, the current Nordic legislation on data protection for countries that are members (Denmark, Finland, and Sweden) is formally based on this directive, and also associated countries (Norway and Iceland) have incorporated this directive as their guiding principle.

However, the directive is open to interpretations and national variation. Therefore, the data retrieval and data protection laws and practices vary greatly across countries. In the Nordic region, the data protection principles are reasonably similar, but still differ to some extent. Rules and vetting procedures are adapted to the fact that large data repositories exist in the form of population based
registers. For this reason, explicit active informed consent has been replaced by different procedures that facilitate research on databases containing a very large number of respondents, making the direct consultation of individuals impossible in practice.

**Ethical vetting**

In principle, all the Nordic countries require that research is subjected to ethical vetting by a legally and subject matter competent committee or government authority.

In some countries, the release of personal data for research in addition requires a separate judgment by the data owner/holder regarding the risks involved for potential harm to the individual. In Sweden, the research is approved according to the Law on Ethical Vetting by regional ethics committees. The risk appraisal associated with the Law on Public Disclosure and Secrecy is done by (each) data owner. Both are needed in order to get access to the research data.

In Sweden and Iceland, the general rule is that ethical vetting is always required when using register data. The ethics board can waive the requirement to consult the data subjects directly to obtain their informed consent, and will often do so if the research is supported by the committee and the data has already been collected in some other context and does not require new contacts with the subjects.

In Norway, the procedure described above is generally used, but there are some exceptions. Strictly anonymized data from a single national register can be used without approval from an ethics committee. A limited set of national registers in Norway have been approved by the parliament and can be used without informed consent. In such cases there is no need for a waiver from an ethics board.

In Denmark, and in Finland in most cases, ethical vetting is not required for studies that only involve the use of already collected government register data, but only if additional data is required. Hence, there is a distinction between the procedures required to be able to use already collected data, and primary collection of data for research purposes. This distinction seems like a reasonable principle to follow.

**Proposal:**

*The group agrees that in the future, one should consider a system where the assessment of risk and the risk of intrusion in the person’s privacy is considered on a magnitude scale rather than assessing the risk on an “all or nothing” basis (i.e. yes or no), and the procedure required is adapted to the estimated “extent of risk” involved.*

**General purpose research registers**

Across all the Nordic countries, primary collection of research data can be performed only with a clearly described and specified research project or program that has been approved. The description provides the basis on which the ethical approval to collect data (and to perform the research in general) is granted. The research has to be described in such concrete terms that it is possible to
evaluate which data are needed. The general policy is to allow collection of or access to already collected data which is limited only to the necessary data needed to perform the research.

For other purposes than research, e.g. official statistics or follow-up of some government activity (e.g. health care), it is possible to collect and store person identifiable data using a purpose that is formulated in more general terms. The construction of general purpose registers solely for research purposes is generally not allowed.

The extent to which the definition of research can be widened to cover several projects and/or be extended in time is not sufficiently clarified. Ethics committees/authorities make judgments on this when they are faced with actual applications. Ethical approvals for general purposes are explicitly prohibited in the Swedish legislation.

Proposal:

Researchers access to existing data repositories and their possibility to build shared repositories for research within general areas of research (e.g. cancer, heart disease, etc.), would be greatly enhanced if research would be considered a legitimate reason to build person identified registries. One way of achieving this is to introduce a national (or cross-national) Research Data Law. The law would establish a general right to create research data registers, describe the allowable principles for collecting the data (including rules on informed consent), and describe the appropriate procedures for researchers to access data from the research registers. Governances defining areas within which research registers are to be established, as well as their content, can be amended to the basic law.

Personal data outside the data owner authority

The release of personal data outside the authorities, for instance to universities and other research institutions, is another problematic area where practices vary across the Nordic countries. In general, in countries where ethical vetting is not required for pure register research there are more restrictive policies for the release of personal data outside the register authority.

Statistics Denmark does not distribute micro level data, regardless of whether they contain person IDs or pseudo numbers, outside of the government authority. Researchers have to visit the authority or get access to micro data online. This makes both domestic research use and collaboration with neighboring countries difficult. Denmark has therefore not been able to participate in some international collaborative analyses using sophisticated methodology that requires individual level anonymous data, or where a deep understanding of the data coding is required.

In other Nordic countries (Sweden, Norway, Iceland, Finland), the statistics authority distributes anonymized micro data, but the policies when it comes to sharing micro level data with other countries in the Nordic region vary. In Sweden, remote access is preferred by Statistics Sweden if at all possible. Physical delivery of micro data with pseudo numbers is the general rule in Sweden for health related research, since there is no online access for the health registries. If necessary for the research, medical researchers may get direct access to person identified data. In general, however, record linkages and other data management operations have to be done in the authorities, (e.g.
Finland, Sweden). The understanding of the data design required in the research may be limited in the authority, which causes delays and errors in the research data. Because of this, the process of obtaining research data is sometimes very time consuming. In Norway, rules for prompt delivery of research data have been implemented.

Proposal:

The formal access procedure is complicated and often unclear to the researcher. In general, it would be very helpful to let a coordinating body take care of all the legal considerations that have to be performed in association with research on personal data. The ideal situation would be to have a single organization to take care of the entire process, acting as a mediator between researchers and data providers.

Review of the European Data Protection Directive

Since 1995 the European Data Protection Directive (95/46/EC) outlines the basic principles of personal data protection in the member states (and associated countries). The current directive constitutes a coherent but not harmonized framework for protection of personal data in the European Union. The national divergences in the protection have caused problems in terms of the free movement of personal data in the Community, and together with the emergence of new data processing technologies, such as federated data bases and cloud computing, it made a revision of the European data protection framework necessary. In November 2010, the Commission published a communication for a comprehensive approach on personal data protection in the European Union. The intention of the review is to achieve increased harmonization of data protection in the EU. The Commission will, among other things,

- examine ways of clarifying and strengthening the rules on consent. The general purpose of this is to increase self-governance. This may entail imposing a more time and resource demanding procedure than the current in obtaining permission to use register data in research. For instance, it may affect the ability of the ethics committee to waive the requirement of personal consultation with research subjects in register studies (e.g. Sweden, Norway). It may also affect the practice to not require ethical vetting for research using only register data that have already been collected before the research (Denmark, Finland).

- clarifying the so-called ‘right to be forgotten’, i.e. the right of individuals to have their data no longer processed and deleted when they are no longer needed for legitimate purposes. This is a possible extension of the rules that sometimes give the research subject the right to have their own data or biological samples destroyed, and may interfere with current rules on archiving research materials.

- consider whether genetic data should be considered as personal data

- further clarify and harmonize the condition allowing for the processing of categories of sensitive data.
• consider new ways to handle data in a secure way, including the use of modern encryption technologies and database handling.

It is expected that the commission will present a proposal for a data protection regulation to the European parliament in the beginning of 2012. The regulation will be directly applicable at the national level. Once the law has been passed, the legal landscape for register based research in the Nordic countries will be changed:

• The first consequence will be a greater harmonization of the rules across the Nordic countries as well as across Europe.

• A second consequence may be that the procedure of obtaining permission to use data will be changed.

• A third consequence may be that the monitoring of personal protection in research by data inspection authorities may take on different forms.

• A fourth consequence is be that the active handling of and archiving of personal data in research may become more difficult.

• A fifth consequence will be that the research community will have to learn and adapt to new data sharing technologies using e-Science applications for distributed databases to a greater extent than today.

A common European data protection regulation could pose both risks and opportunities for register based research in the Nordic countries. This is why it is imperative that the research community engages in the parliamentary examination process of the new regulation as well as at the adoption stage of the regulation. The aim would be to improve the conditions for register based research in each country and for data sharing across borders using the new common legal framework.

Proposal:

The group suggests that the NCDH as one of its main tasks should investigate how to implement the new European data protection legislation and the legal changes that will be the consequence of this in each of the Nordic countries, thereby creating a proposed common policy that is beneficial to register based research. This will require resources in terms of legally competent personnel in addition to the basic staff for the NCDH in each Nordic country.
E-science solutions - Technical development of a distributed data infrastructure

The requirements on data access involve combining data from large and sensitive registers that cannot be physically moved. Sometimes data sharing requires obtaining data from many heterogeneous sources, such as sample survey data, medical treatment records from hospitals and clinics, social and demographic population data, bio bank sample data, and centralized health registries. Both from a legal, ethical, and practical point of view, it is preferable if these data are not moved or duplicated into many copies. How to develop e-Science solutions to these problems have not been considered by the group, but the potential of this possibility is substantial in the long run, especially in light of the common legal framework that may be the result of the ongoing data protection revision.

In a distributed system, the data structure needs to be reasonably compatible such that researchers can understand how to combine data from various sources. The documentation needs to be accessible through an easy-to-reach common portal. The data should be configured such that it is possible to link data across organizations and data sources. The possibility to use cloud computing in association with distributed data sources should be explored.

The use of trusted third parties (TTP's) to handle research data bases, the use of federated techniques in connecting large data holdings with each other, the use of cloud computing, the use of encryption techniques, and remote access to data will develop in the future, and should be considered in the Nordic collaboration.

There have been several efforts to pool data in pan-European projects; one of the largest being the European Prospective Investigation into Cancer and Nutrition (EPIC). Another example is the pan-European collaboration GenomEUtwin under the 5th Framework Programme, using the first project driven hub and spoke network with federated databases. While these projects have yielded scientific results, none of them has created a sustainable e-infrastructure. Besides these initiatives, harmonisation of biomedical data has been addressed in collaborations like the Smart Open Services for European Patients epSOS. In Sweden, the Bio bank Information Management System (BIMS) is the most developed solution for these harmonisation problems, using federated databases. BioGrid in Australia is a federated system with many users, but it only supports one software solution (SAS). In the Netherlands, the String-of-Pearls (PSI) initiative aims to develop a centralised infrastructure for prospective studies among the 8 university medical centres. The current Swedish structure for data sharing consists of a fast virtual net for the distribution of data (Swedish University Net, SUNET) and a system for providing technical resources for computing and storage of data (Swedish National Infrastructure for Computing, SNIC). Similar technical structures exist in other Nordic countries, for example within the Finnish Center for Scientific Computing (CSC). However, the corresponding
Nordic national or joint infrastructures for safe database handling, retrieval and analysis of data that cover various data needs do not yet exist.

**Data warehouse solutions**

Systems for remote access to a centralized warehouse, including analysis capacity, exist in Statistics Sweden and Statistics Denmark. Similar systems are being developed in Norway and Finland. This works primarily for finished data collections of moderate sizes. For most of the data coordination needs using large and/or rapidly changing data holdings, the centralized data warehouse solutions are not feasible of legally possible.

There is also an ongoing discussion about the possibilities of centralized solutions with remote access to adapt to the heterogeneous needs of different research groups in terms of data access and specialized software for data analysis. It is particularly important to develop technology that will preserve the same flexibility and complete data access that exists when the data is delivered physically to the researcher. Currently, researchers using remote access have experienced problems in both these areas, which in many cases have led to the abolition of the remote access principle in practice. The migration from physical data delivery to remote access must therefore be gradual, and the experience from current remote access systems used in their future technical development. Several legal problems, such as ownership issues and the archiving of remote access research data, also remain to be solved.

**Federated and cloud data environments**

In a federated system, the goal is to provide data access solutions for coordination for large and complex data sets without physical shipping of the original data. The main goal is to provide a system that can handle person identified sensitive data efficiently and in a legally acceptable way. In the creation of such systems, it is possible to build on the knowledge of federated databases listed above. The database field is moving fast, so it is also necessary to explore which services from a federated system are also suitable for cloud computing.

**Database development**

Independently of the development of a remote retrieval and analysis system is the development of recommendations and tools for handling the data in a single database. A heterogeneous landscape of data holdings needs to be transformed into mutually compatible formats without requiring too much of a centralized structure. This would mean that the development of software tools for the combination of different database formats needs to be developed. The goal for this work is to provide a minimum set of common data definitions such that users and data owners can communicate in an efficient way. Common or compatible data formats are independent of the development of data access methods, and provided by the parallel development of software tools and support systems for users in transferring data into acceptable formats for exchange.