Reinforced Nordic collaboration on data resources

Challenges from six perspectives

Reinforced Nordic collaboration on data resources

Challenges from six perspectives

Marie Sandberg, CSC

Reinforced Nordic collaboration on data resources
Challenges from six perspectives
Marie Sandberg, CSC

TemaNord 2012:514
ISBN 978-92-893-2342-0

© Nordic Council of Ministers

Print: Kailow Express ApS

Copies: 80

Printed in Denmark

This publication has been published with financial support by the Nordic Council of Ministers. However, the contents of this publication do not necessarily reflect the views, policies or recommendations of the Nordic Council of Ministers.

www.norden.org/en/publications

Nordic co-operation

Nordic co-operation is one of the world's most extensive forms of regional collaboration, involving Denmark, Finland, Iceland, Norway, Sweden, and the Faroe Islands, Greenland, and Åland.

Nordic co-operation has firm traditions in politics, the economy, and culture. It plays an important role in European and international collaboration, and aims at creating a strong Nordic community in a strong Europe.

Nordic co-operation seeks to safeguard Nordic and regional interests and principles in the global community. Common Nordic values help the region solidify its position as one of the world's most innovative and competitive.

Nordic Council of Ministers

Ved Stranden 18
1061 Copenhagen K
Phone (+45) 3396 0200

www.norden.org

Content

Summary	7
1. Introduction.....	9
2. Background.....	11
3. Scope of the study.....	13
4. Nordic collaboration.....	15
4.1 Register-based research.....	16
4.2 Six perspectives.....	19
4.3 Summary of challenges.....	30
5. National progress.....	33
5.1 Denmark.....	35
5.2 Finland.....	36
5.3 Iceland.....	38
5.4 Norway.....	39
5.5 Sweden.....	41
6. International view	43
7. References.....	47
8. Sammanfattning	49

Summary

This study is mapping the most significant challenges and obstacles for a reinforced Nordic cooperation on data resources. Focus is put on existing national databases and registers established mainly for administrative purposes but also the question of newly-generated scientific data is handled. The challenges are analysed from political, legal, ethical, organisational, technical and financial perspectives. The broad scope targets primarily policy makers involved in eScience development on national and/or Nordic level. Involved parties in the study are Nordic Council of Ministers, NordForsk, Finnish Ministry of Education and Culture and CSC – IT Center for Science.

Special focus has been put on the challenges for register-based research, since all the Nordic countries have a vast amount of population-based registers which are considered not to be used to their full potential in various research fields. One of the challenges is to find way to combine these registers with other research data and the Nordic countries have to find their own way of unique collaboration methods, since there are no equivalent pre-conditions for this kind of research in the rest of Europe.

The main challenges found from an *organisational perspective* are the scattered data resources, meaning both the high number of different databases/registers as well as the wide range of organisations involved in maintaining these data resources. Other organisational challenges are the *lack of research perspective in authorities* as well as the *lack of a common language between researchers and employees of the authorities*. For register-based research the *difficulty to access personal data* is a specific organisational bottleneck. There is also a challenge when it comes to newly-created research data regarding *researchers own hesitation towards data sharing*, despite existing supporting data infrastructures. *The lack of Nordic perspective in national strategies and initiatives targeting research data resources* also poses a challenge for reinforced Nordic cooperation, since there is a risk of developing policy directives which are not aligned on a Nordic level.

From a *legal perspective* the main challenges found are *different legislation regarding mainly access to personal data in the Nordic countries and between different authorities*. There is also an insecurity regarding the *revision of the EU Data protection legislation* and how it will affect the national legislation and possibilities for register-based research, even though a new EU regulation with requirements of revising the national data protection laws, also could give a necessary momentum for the

Nordic countries to look into further harmonisation of legislation on a Nordic level.

The *financial perspective* highlights the benefits of *cost-efficient research*. A huge amount of public funds are spent on maintaining existing public registers and databases as well as gathering new data and it is highly cost efficient to use these resources to their full potential in research. A selection process needs to take place to *determine which databases and registers are worth the costs of improving and making truly sustainable* by integrating them into a reinforced Nordic database infrastructure. To promote newly-generated scientific data sharing *funding incentives* may be used. Data sharing could be promoted through a common Nordic system where research projects and researchers receive funding credits for sharing data. From a financial perspective it is important with *pooling of national funds for Nordic research benefit*; national funds should be set aside to ensure Nordic data resources collaboration. One solution could be to fund for instance ESFRI projects supporting sharing of data resources, but to promote a Nordic perspective within these EU initiatives – a second layer of collaboration inside the bigger collaboration.

The *ethical perspective* mainly handles register-based research and biobanks where the challenge is to use the advantage of the unique Nordic conditions for register-based research without violating the personal privacy and integrity. One important issue is to keep *the societal trust and an open ethical debate* towards this type of research, not only on a national level but also on a Nordic level by for instance making room for more Nordic channels and platforms for public dialogue and involvement. *The role of national Ethics Advisory Boards and Ethics committees* in a reinforced Nordic cooperation also have to be clarified and the working methods need to be aligned.

The *technical perspective* deals mainly with the challenge of *interoperability and how to go from pilot project to large-scale solutions*. The technical perspective also underlines *the importance of political support to develop technical solutions and the monitoring and use of best practices of technical solutions in international initiatives and projects*.

The study also gives a national overview of the current progress in the five Nordic countries, to raise awareness of important national initiatives which can contribute to a stronger collaboration on the Nordic level. The final chapter gives example of current interesting initiatives outside the Nordic countries dealing with collaboration on data resources, , such as EUDAT and Data Without Boundaries. In some of these initiatives Nordic countries are taking part but as separate countries, rather than with a Nordic perspective.

1. Introduction

In 2006, the Nordic Council of Ministers formed an adhoc eScience Work Group assigned to propose a joint Nordic eScience strategy. The Work Group delivered the strategy in July 2007, emphasising the need for long-term, sustainable collaborations within Nordic eScience infrastructures and proposing new Nordic eScience programmes within research, infrastructure and education. This initiative was taken further by the research councils of Denmark, Finland, Sweden, Iceland and Norway under NordForsk's NORIA-net instrument, forming The Nordic eScience Initiative (eNORIA). In the course of their work, the participants of eNORIA were asked by the Nordic Ministers for Education and Research to propose an action plan to realise the Nordic eScience strategy. The action plan, consisting of 10 concrete action points was elaborated in 2008.

In 2009 it was decided to turn eScience into a Nordic "globalisation initiative", meaning that it would become one of the main priorities for the Nordic Council of Ministers. NordForsk has been given the responsibility of coordinating the implementation of this initiative during 2011–2015, with a thematic focus on climate change and research on health/social preconditions for health.

Finland had the Presidency of the Nordic Council of Ministers in 2011 and organised an eScience conference on 29–30 September 2011 in Helsinki¹ to put further focus on the subject. During the preparations of the conference, the scope of this study was initiated with the objective to make an inventory of the main challenges for a reinforced Nordic cooperation regarding the use of data resources, in particular public databases and registers and to present the first findings at the Helsinki eScience conference.

Since reinforced cooperation regarding data resources is to be solved not only at the level of authorities, it was decided to give the study a broader scope, where for instance the political and ethical dimensions are considered. There are already many excellent national studies focusing on separate barriers, such as technical and legal challenges and there is a need for an overall inventory of the challenges to be brought up and solved at policy level. Hopefully, the study can serve as an interlink between different problem areas, which can be of support for the further work of the Nordic eScience globalisation initiative, coordinated by NordForsk.

¹ <http://www.escience.fi>

Involved parties in the study are Nordic Council of Ministers, NordForsk, Finnish Ministry of Education and Culture and CSC – IT Center for Science. The project leader made use of a reference group consisting of representatives from CSC and NordForsk and a special thank you goes to Maria Nilsson PhD, Senior Adviser, NordForsk and Pirjo-Leena Forström, Service Area Director data services for science and culture, CSC.

The main method has been desk research and focus is put on summarising and analysing already existing research, reports and strategies in relevant areas. Some interviews have also been carried out to complement the written material and especially the following persons contributed with clarifying and interesting discussions: Juni Palmgren, Secretary General within research infrastructures Swedish Research Council, Gudmund Høst, Special Adviser Research Council of Norway and Chair of the eInfrastructure Reflection Group (e-IRG), Marjut Salokannel, Research Director Faculty of Law, University of Helsinki and and Leif Laaksonen Director collaboration, CSC.

2. Background

The Nordic eScience strategy² – the strategy for the Nordic Council of Ministers from 2007 states the importance for cutting edge research of establishing a Nordic infrastructure for databases and data repositories, due to the Nordic unique registers and databasess.

The eScience action plan³ developed in 2008 by eNoria to realise the Nordic eScience strategy clearly states the importance of utilizing the rich source of databases and reporsitories in the Nordic countries. Action 7 “*Open access to National Data Repositories*” underlines legal, organisational and policy restrains rather than technical issues to enable the combined use of national databases and repositories within selected research areas:

“The Nordic countries have comprehensive databases and data repositories owned by national authorities for administrative or monitoring purposes. No other part of the world has such comprehensive data, namely within the area of social sciences, the humanities. In addition, new valuable scientific data is collected and generated all the time – at an almost explosive rate. Most of these unique data structures are currently underused for research – more often than not, due to legal, organisational and policy restraints rather than technical. An important goal for the infrastructure action line is to enable the combined use of the existing national databases and data repositories within selected research areas, as well as to initiate activities improving the possibility of using newly-generated scientific data. In so doing, the many non-technical restraints must be overcome. The lack of documentation calibration and validation of data records has been found problematic by both the OECD³ and the Council of the European Union, which have released recommendations on data access and preservation.” (p. 13)

The report “Knowledge Infrastructure for the Fifth Freedom in the Baltic Sea Area”⁴ by the Nordic Council from 2010 also states the importance of a Nordic database infrastructure and mentions the organisational aspect apart from necessary technical solutions;

² http://www.vr.se/download/18.6f6ddf01189d73e59380002244/nordic_escience.pdf

³ <http://www.nordforsk.org/en/publikasjoner/nordic-escience-action-plan/view>

⁴ <http://www.norden.org/en/publications/publikationer/2010-516>

“In the Nordic countries the authorities have built up unique comprehensive databases for administrative and other purposes that would be unparalleled on the international arena if effectively made available for research. In addition to providing technical solutions, sustainable organisations for providing scientific databases are important” “In focus for national and joint Nordic initiatives, key to effectively contribute to the European ecosystem of digital repositories, are technical solutions for federated databases, policy frameworks for harmonization of data and interfaces, as well as administrative and legal issues” (p. 23–24)

The report also recommends a small number of pilot projects, using cross-border resources to start developing necessary technology and policies which facilitates cross-border resource-sharing.

There is in other words a Nordic awareness of the advantages of a reinforced Nordic cooperation regarding data resources, as well as an understanding that there are many areas, other than the pure technical aspects that must be coordinated to fully explore these advantages. In the Nordic eScience Strategy from 2007, it is stated that the work on establishing databases and data repositories as national infrastructures is in an early phase in the Nordic countries. In the nearly five years that have passed since the development of the strategy, further initiatives have been taken both on the national level as well as the Nordic level. The activities of NordForsk are very good examples of a joint initiative in this area, but there are still many initiatives on the national levels that are not necessarily shared between the countries. The information flow from the national to the Nordic level could still need more coordination and formalised routines, since this is an area under very fast development.

3. Scope of the study

The study is mapping the most significant challenges and obstacles for a reinforced Nordic cooperation regarding data resources, in particular

- the use of existing national databases and registers owned by national authorities for administrative or monitoring purposes
- the use of newly-generated scientific data

The challenges are analysed from political, legal, ethical, organisational, technical and financial perspectives.

Special focus has been put on the challenges for register-based research. Population-based registers built up by authorities in combination with the system of social security numbers is a unique feature of the Nordic countries and are often referred to as a “gold mine” from a research perspective. The challenge is to find way to combine these registers with other research data and the Nordic countries have to find their own way of unique collaboration methods, since there are no equivalent pre-conditions for this kind of research in the rest of Europe. Although the Nordic collaboration regarding some registers, such as the cancer registers and health care registers has come a long way, there are still many research fields where the registers are considered not to be used to its full potential.

There is no need to put too much focus on the benefits related to a reinforced Nordic cooperation, since this is a clear starting-point for the study. The objective is to make an “inventory” of the challenges and problems and focus is put on the overall picture, rather than detailed information on separate challenges. It is not in the scope of the study to find solutions for the identified challenges and problems, but some suggestions are given for certain problem areas. Hopefully, the study can pave the way for future solutions by painting the big picture, thus discovering possible interlinks between different problem areas. The broad scope targets primarily policy makers involved in eScience development on national and/or Nordic level. For this reason, the political perspective is a “meta perspective” which is present in the description of all the different challenges. Research representatives and policy representatives sometimes have difficulties in communicating and understanding each other and for decision-making purposes it is necessary with an overall political understanding of the impact of these challenges rather than in-depth, detailed knowledge.

The study also gives a national overview of the current development in the five Nordic countries, to raise awareness of important national initiatives which can contribute to a stronger collaboration on the Nordic level.

The final chapter gives example of current interesting initiatives outside the Nordic countries, dealing with collaboration regarding data resources. In some of these initiatives Nordic countries are taking part. It is advisable to make use of best practices from these initiatives and adapt it to the specific Nordic conditions if necessary. The Nordic countries are taking part in some of these initiatives but as separate countries, rather than with a Nordic perspective.

4. Nordic collaboration

Collaboration regarding data resources origins from the basic idea that the value of information increases when it is shared. There is a massive amount of information stored in databases and registers owned by public authorities in the Nordic countries. These databases have been built up mainly for monitoring and administrative purposes but can be used for research purposes to a much greater extent than today. Population-based registers are very important examples since they are so unique for the Nordic countries. But there are many other variants of databases owned by public authorities, which do not contain so sensitive information as population-based registers, for instance data on climate and environment.

It is however not possible to only take these public authorities data resources into account for a reinforced Nordic collaboration. There are also a massive number of data resources with newly-generated scientific data, specifically designed for the purpose to serve research, which also need a stronger Nordic collaboration. It is the possibility to use and combine data from all these different data resources that will create truly groundbreaking research possibilities across disciplines and across national borders in the Nordic region.

Environment Climate Data Sweden

One current example of a national service where these different data sources comes together is the Environment Climate Data Sweden (ECDS). ECDS is funded by the Swedish Research Council and is a service facilitating the searching, publication and long-term accessibility of data for research in the fields of environment and climate. This can lead to new ideas of how to use and combine data from different sources and also provide opportunities to find new collaborators in other disciplines. The data portal includes metadata for datasets and links or references to the data sources. At the moment ECDS deals mainly with data provided by researchers but there is also some data from the Swedish Meteorological and Hydrological Institute (SMHI). SMHI is currently building a parallel data portal which will be accessed from the ECDS portal. When data from other public authorities become increasingly available, it will also be included in the portal.

More information: <http://www.smhi.se/ecds>

4.1 Register-based research

Register-based research is research where population-based register data is used, either solely or in combination with other data sources. All the Nordic countries have a vast amount of population-based registers. It is unusual in the rest of Europe, where much of the gathered information about for instance health is survey based. The Nordic register data is often referred to as a gold mine for various scientific fields. Registers for administrative and health purposes were established in the Nordic countries already in the 1950s. Since then, the number of registers have grown to include for instance cancer registers, occupational diseases, cause of death, register of dental implants, register on social assistance, child welfare, pensions, income and taxation to mention a few. Register-based research exists in various sciences, such as clinical medicine, health care and sociology. The Nordic system with a social security number for all citizens is the key for register-based research, with enormous possibilities for cross-sectoral research. Some of these facilities are present in other countries, but the combination of population based registers and the social security number used for such a long period of time is unique for the Nordic countries. The challenge lies in how to strengthen the infrastructure for register-based research, especially since most registers have not been gathered for research purposes. One of the most frequent barriers mentioned is the notion that registers are not used or matched with other data to their full potential.

It is possible to obtain data for research from most registers, but the researchers need to apply for permission or get a statement from ethical board or data protection body. The method varies in the Nordic countries. Also the process of combining/matching data varies in the Nordic countries. It can either be researchers themselves or the authorities that does the work. Informed consent is in general not necessary from individuals included in a register study. There is however strict rules for published information to prevent it from being linked and traced back to individuals (cryptation etc).

4.1.1 *Registers and Biobanks*

The Nordic biobanks are excellent examples of the possibilities to combine population-based registers with other data and an area where Nordic countries can take the lead at EU and international level. National biobanks collect data on a large scale on biological and environmental/lifestyle characteristics of individuals (data on health, life habits and

genetic profiles etc). Using data from biobanks in combination with other registers, researchers try to identify the genetic and environmental factors underlying diseases, to improve prevention, diagnosis and treatment. Collecting biological information from people with illnesses has a long history, but collecting data from healthy people is relatively new and key to biobanks. There is an on-going debate regarding the ethical challenges to match more and more data sources since it can be hard to estimate the consequences. When does the research risk becoming a threat to personal integrity and privacy?

BBMRI Nordic – a Nordic collaboration example

At the European level, BBMRI (Biobanking and Biomolecular Resources Research Infrastructure) is one of the ESFRI priorities.

To work towards a harmonisation of the biobanking process on the Nordic level, NordForsk funded the Nordic network BBMRI Nordic in April 2010. This initiative would also serve as a common Nordic biobanking voice on the international research infrastructure scene and the network will bring Nordic expertise and support into the European BBMRI initiative.

BBMRI Nordic recently established a joint Nordic biobank-based research project called “Joint Nordic Biobank Research Infrastructure”. A harmonisation process on a Nordic level will enable studies where data and samples from many different Nordic countries can be pooled and combined. Colorectal cancer has been chosen as the medical-based pilot use-case, since large-scale patient material is needed in this research area. The suggested pilot project is divided into three phases of which NordForsk is funding the first two that will run 2011–2013. The third phase may include the establishment of a legal entity to facilitate the creation of Nordic Biobank registries.

More information: <http://www.bbmri.se/en/About-BBMRIse/BBMRI-Nordic/>

Nordic people share many values which also facilitates research co-operation in this aspect. They are for instance most positive to Biobanks and also in general positive to share this data internationally according to the Eurobarometer 2010,⁵ which is shown on the following pages. The positive attitudes make ethical constraints less problematic, since there is a shared value in the societies of the benefits of this type of research.

⁵ Eurobarometer 73.1 “Europeans and biotechnology in 2010”,
http://ec.europa.eu/public_opinion/archives/ebs/ebs_341_winds_en.pdf

Figur 26: Would you be willing to provide information about yourself to a biobank?

% respondents

Figure 31: Support for sharing and exchange of personal data and biological materials

% respondents

4.2 Six perspectives

In the following chapter, the challenges and opportunities connected to reinforced cooperation regarding data resources will be described using six broad perspectives presented below. The *political perspective*; issues that arise on the political level is a “meta perspective” and does not have a separate section. Instead it will be present in all the other perspectives, since the study is mainly targeting eScience policy-makers with the aim of Nordic reinforced collaboration, which requires an overall political understanding of the impact of all these challenges.

The organizational perspective: various challenges and bottlenecks which occur on the organisational level with implications for the Nordic collaboration on data resources.

The legal perspective: laws and regulations on national and EU level of importance for the Nordic collaboration.

The financial perspective: how can reinforced Nordic database collaboration be motivated from a financial perspective? What are the economical challenges?

The ethical perspective: challenges connected mainly to our moral and ethical framework, often linked to the specific features of our Nordic cultures and societies. Register-based research and biobanks are the main topics for this perspective.

The technical perspective: challenges which have to be solved mainly on the technical level to support reinforced Nordic collaboration on data resources.

4.2.1 The organizational perspective

The organisational perspective describes some of the main challenges on the organizational level which are of importance for Nordic database and register collaboration. This perspective contains many different aspects depending on the variety of databases; the challenges are very different for population-based registers and other databases in public authorities compared to databases with newly created scientific data.

Scattered data resources

One well-known challenge for a functioning database infrastructure is the issue of quantity, which refers to both the high number of different databases/registers as well as the wide range of organisations involved in maintaining these databases. No Nordic country seem to have a complete overview of registers and databases that exists in various authorities. The scattered picture makes it harder for the researcher to find or even be aware of the existence of useful data. To get a full picture of all the Nordic databases and registers may be an overwhelming task. One solution to start with would be to map all the databases and registers related to the research themes climate change and health, which are prioritised in the Nordic “globalisation initiative”.

Another problem due to the range of organisations involved in maintaining these databases is the different praxis of data management which exists in different authorities. Increased data management harmonization and coordination would facilitate data sharing and the use of multiple data sources. The challenge is that it will require extensive work from authorities to change their data management routines. Policy frameworks to support this work have to be developed in close collaboration between decision-makers and the authorities.

Currently, the Ministry of Finance in Finland is reinforcing sectoral inventories of data produced by public sector organisations. The resulting data will be put in a metadata catalogue. To create Nordic interoperability of these metadata catalogues could also be a way to reinforce the Nordic collaboration.

Authorities lacking research perspective

Since many databases and registers have been developed originally mainly for administrative or monitoring purposes they are not designed for research and to assist in research is not the main task of the authorities. Data documentation, production, storage etc have often been developed without taking the research perspective into account.

The lack of common language

The lack of a common language refers mainly to the difficulties for researchers and employees in authorities to understand each other's work. It is closely linked to the challenge above; authorities lacking research perspective. The authority may not have the staff expertise required to understand the researchers' data requests. But it also goes the opposite way; researchers lack the perspective of authorities and do not understand how to fill in requested forms for or that there are many other important tasks for the authority to deal with than assisting the researcher. Lack of common standardized concepts and vocabularies is a big challenge not only between researchers and administrators but also for people involved at the ICT-level of the authorities.

The lack of common language can of course also refer to barriers between the different Nordic languages, since public databases and registers are meant for national purposes. Databases with newly-generated scientific data often have English as working language, which also helps researchers outside the Nordic countries. The European Research Infrastructure for Biobanking and Biological Resources have tried to solve the language barriers and the need for joint understanding of various definitions by creating a "biobank lexicon" with (currently) seven different languages and basic datasets in various languages. As the data itself very often is just numbers, it is helpful to use standardized vocabularies for data description, which usually have simultaneous translated versions and are easy to share.

Other organisational challenges specifically for register-based research

In register-based research, which contains mainly confidential information, the researcher has to apply for data, often from many different authorities. These authorities follow different practices and laws regarding the delivery of data. The authorities have to handle the research applications, often ask for additional information and then do the actual compilation of the data sets. The time that is passing from permit appli-

cations to final delivery of data sets can cause serious delay in the research projects and the data can even be outdated. The register descriptions are often unclear and the content and storage format of the metadata is not standardized.

Researchers hesitation towards data sharing

Organisational challenges for the use of newly-generated scientific data are researchers hesitation towards data sharing despite existing and supporting infrastructure. Researchers' behavior is also part of the organizational perspective, since researchers are influenced by their organizational culture. The challenge lies not only in establishing an infrastructure which supports data sharing but to create an organizational culture of data sharing. There are differences between scientific disciplines. In a newly published EU survey⁶ on open access in the seventh framework programme (FP7), project coordinators were asked how they would view an open access to data in their research area, providing that all relevant aspects such as ethics, confidentiality and intellectual property have been considered and addressed. It showed that researchers in environment would be more positive to share their data than researchers dealing with projects targeting health or energy. Researchers' hesitation to share data *despite existing infrastructures* is also found to be one of the main challenges for data sharing according to the EU project "Opportunities for Data Exchange" (ODE) which is described more in the last chapter. Unless data sharing is seen as an actual added value by the researchers it will be hard to make them share their data voluntarily.

Lack of Nordic perspective in national strategies targeting research data resources

The described organisational bottlenecks and challenges are well documented and have been highlighted in a number of reports and strategies over the past few years. The Nordic countries are working to solve these issues on the national level, which is described more in the national overview chapter. In Finland, a cross-sector coordination group is currently working to enhance data issues from a research perspective and plan legislative reforms that will enhance the wider utilization of data resources. In Iceland a task group for future arrangement of databases is currently being established, which deals with database coordination, open access, administration and maintenance of databases. An open access committee in Denmark recommends national planning of free access to as well as long-term preservation of primary research data by establishing an interdisciplinary cooperation organisation, and the Database Infrastructure

⁶ http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-open-access-in-fp7_en.pdf

Committee (DISC) in Sweden has been working for a few years together with the Swedish National Data Service on co-ordinating existing and new research databases. The Swedish Research council has also recently produced a new guide for research infrastructure⁷ where improved conditions for register-based research is highlighted.

Many technical and organisational challenges for register-based research are dealt with in projects such as Finnish Midras – Micro Data Remote Access System, RAIRD in Norway – Remote Access Infrastructure for Register Data, and SIMSAM – Swedish Initiative for research on Microdata in the Social and Medical Sciences.

Even though these strategies and initiatives indicate that something is happening regarding improved database and register infrastructure, the challenge still lies in the lack of a sufficient Nordic perspective. NordForsk has a clear mandate to promote Nordic research policy and collaboration but it also needs to be reflected and prioritised when developing the national strategies and policy directives. How come that the Nordic cooperation is not described and targeted in a similar manner in the national strategies as for instance the ESFRI collaboration?

Already in the Nordic eScience strategy from 2007 it was pointed out that Nordic cooperation on these issues is urgently needed, before the national database infrastructure solutions are in place. Harmonised database infrastructure on the Nordic level should be developed jointly with the national initiatives, both to learn from best national practices and prevent ending up with very different national systems and solutions which would in worst case risk to complicate the Nordic collaboration.

4.2.2 The legal perspective

The current 1995 EU data protection directive lays down certain goals for the processing of personal data which must be achieved in every Member State. National authorities had to adapt their laws to the data protection directive, but could still be flexible in how to meet these goals. This has resulted in very different national practices regarding the handling of research data, also when comparing the Nordic countries.

In Denmark, Norway and Sweden processing of sensitive data for research purposes needs approval by the Data Inspection Agency (although in Sweden, the approval of the National Data Inspection Agency is not necessary if a research committee has approved the processing.) According to the legislation in Denmark, Iceland, Norway and Sweden, statistical data may be released with identification data for scientific research and statistical surveys. In Finland, the current statistics act

⁷ <http://www.vr.se/download/18.9b19d771342fcd4fc8000817/Vetenskapsradets+guide+till+infrastrukturen+2012.pdf>

(which is about to change) makes it very difficult to obtain micro-data for research purposes. One condition in all countries is that access to confidential data for statistical or research purposes must not cause any damage or be detrimental to the data subjects. Both Denmark and Sweden has developed online data-base systems to obtain micro data for research purposes.

The EU Data protection legislation is currently under revision and the Commission published a proposal for a data protection regulation on 25th of January 2012.⁸ A regulation is directly applicable in all Member States and cannot be changed to fit certain national requirements. In the proposal (which may still be subject to change), there is a requirement for explicit consent from the concerned individual before their data is used. This has created concern regarding the continued use of population-wide person-identifiable registries in the Nordic region, and how this would affect register-based research. However, according to the latest proposal, register data collected for *other purposes than research* may be used for research purposes as long as it is anonymised or encrypted. This prerequisite covers basically all the Nordic population-based registers.

The changes in the EU data protection directive could be seen as a potential threat for the Nordic countries but could also be seen as a chance to further reinforce the Nordic cooperation. Once the Data Protection regulation comes into force, the Nordic countries will have to revise their data protection laws. The required major changes in the laws could possibly give a momentum to look into a harmonisation of the Nordic legislation and maybe even provide an opportunity for setting up a common Nordic federated database system.

Not only on the Nordic level but also on the national level, a number of legal aspects complicate the construction of a research infrastructure in the form of research databases. An excellent report on the legal conditions for a database infrastructure in Sweden has been produced by Swedish research council in 2010; "Rättsliga förutsättningar för en databasinfrastruktur för forskning".⁹ Various legal aspects such as the Personal Data Act, the Ethical Review Act, The Public Access to Information and Secrecy Act is analysed with the general recommendation to review the laws and regulations and weigh in the opportunities and needs of research in relation to other interests such as the individual's right to privacy. The other Nordic countries may benefit from carry out similar analysis, in parallel with the work on harmonisation of legislation on the Nordic and EU level.

⁸ http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

⁹ <http://www.vr.se/download/18.28ca860b1293e247fc480006831/Report+11+2010.pdf>

4.2.3 *The financial perspective*

How can reinforced Nordic database collaboration be motivated from a financial perspective? What are the economical challenges?

Cost-efficient research

A huge amount of public money is spent on maintaining existing public registers and databases as well as gathering new data for monitoring and administrative purposes. It is highly cost efficient to use these resources to their full potential also for research, especially considering for how long period of time data has been gathered in many cases. To extend this efficient use of public data for research purposes from the national level to the Nordic level is even more cost efficient, considering the fact that workload and costs can be shared in bigger research projects. Newly created scientific data could also be reused and shared in a more cost-efficient manner, especially considering researchers own hesitation to share their data despite existing supporting infrastructure, as mentioned under the organisational perspective.

Not all databases can survive

One financial challenge is however that it would be costly to spend money to harmonise, improve and organise *all* existing databases and registers. A selection process needs to take place to determine which existing data resources are worth the cost of making truly sustainable by integrating them into a reinforced Nordic collaboration. This selection process would benefit from being done as a joint Nordic exercise instead of separate selection processes in the different countries. As a bonus, the exercise would boost the common awareness and knowledge of different data resources in the Nordic region.

Funding incentives for data sharing

Efficient data sharing also minimises the risk of different researchers producing the same scientific data without being aware of it – which also is a waste of public money. To promote newly-generated scientific data sharing it can be useful to use funding incentives. Data sharing should be promoted through a common Nordic system where research projects and researchers receive funding credits for sharing data.

Pooling of national funds for Nordic research benefit

The Nordic countries have a strong tradition for collaboration but specific funding for Nordic co-operation is not always adequate. In contrast, collaboration among other European countries is becoming increasingly important and is often better resourced. National funds should be set aside to ensure also the Nordic database collaboration. More and more international collaboration which is necessary in today's research takes however a big piece of the national research infrastructure budgets. One

solution could be to fund for instance ESFRI projects supporting database infrastructures, but to promote a Nordic perspective within these EU initiatives – a second layer of collaboration inside the bigger collaboration.

4.2.4 *The ethical perspective*

One of the main ethic challenges regarding databases and registers is the handling of data which contains personal information, which especially concerns person data used in register-based research. There are ongoing initiatives in all the Nordic countries to promote the use of registers for research purposes. The main ethical challenge in this respect is to be able to take advantage of these unique Nordic conditions, without violating the personal privacy and integrity.

In particular biobanks has been the subject of ethical debate. Critics raise questions about privacy, confidentiality and commercial interests regarding the biobanks and about who is going to regulate them. The more and more complex combinations with other data, such as registers has also caused some concern on how to be able to estimate the possible consequences for the personal integrity of such research.

The importance of societal trust and an open ethical debate

As shown in the previous chapter, in a European context, the Nordic countries are most positive to Biobanks and also in general positive to share this data internationally. The positive attitudes show a shared value in the societies of the benefits of this type of research, which is one important factor to consider from an ethical perspective.

Even though there may be a current societal trust in the Nordic countries for biobanks and research involving sensitive data, this is not constant state. Every now and then incidents and challenges occur, which makes the public trust fluctuate. These ethical challenges are much more likely to be solved with an open and ongoing ethical debate.

A recent example of a debate with ethical implications is the decision of the Swedish Data Inspection Board that the ongoing collection of biological samples and health data to the large biobank “LifeGene” is in breach of the Personal Data Act. Karolinska Institutet that runs LifeGene must stop collecting further data and is not allowed use already collected data. The reason for the decision is the general purpose of LifeGene which is “future research”. According to the Data Inspection Board this purpose is too vague for participants to give consent to, even though they participate voluntarily. LifeGene is not a research project, it is intended as infrastructure of future biobank projects and this does not comply with current legislation or ethical principles according to the Swedish Data Inspection Board. In its decision, the Data Inspection Board writes that the database management of personal data for future research is such an important issue that the parliament and government should review current legislation. The case is widely debated in the Swe-

dish press and Karolinska Institutet has appealed the Data Inspection Boards ruling.

Another example is the Health Sector Database experience on Iceland, which resulted in an extensive public debate which included many ethical aspects. During the late 1990s, the large research company deCODE genetics attempted to set up an Icelandic Health Sector Database (HSD), a centralised database of personally non-identifiable health data.¹⁰ In 1998 the Act on Health Sector Database was passed which gave the company access to the national health record data without the expressed consent of individuals. In addition, information on deceased individuals would be automatically included, despite the potential privacy interests of relatives and individuals. Although the idea to allow the company access to the national health record data without consent of individuals was never realised, it caused an intense discussion, mostly around the questions of consent, privacy and individual rights.¹¹ The ethical debate in the end resulted in change of legislation; prior consent of the patient on Iceland is now required to be part of any database, in which health data on individuals are permanently retained for research purposes. According to the Europabarometer from 2010, Iceland is the Nordic country which is clearly most positive to Biobanks and to share their data internationally. This may be the result of the wide public debate which was going on for so long.

The challenges for Nordic reinforced collaboration would be to lift the ethical debate and societal trust to a Nordic level. One way to secure public engagement and involvement could be to “empower the citizen”, which is described in the book *Bioethics, Politics and Business*,¹² based on seminars held by the Nordic Committee on Bioethics. One of suggestion is to try to strengthen professional media and scientific journalism, which provides the citizens with information and critical ethical analysis. Another suggestion is to make room for more Nordic channels and platforms for public dialogue and involvement.

The role of national Ethics Advisory Boards

The Nordic countries have national bodies responsible for research ethics in general, such as the National Advisory Board on Research Ethics in Finland, the Central Ethical Review Board in Sweden and Danish Council of Ethics. These boards (with some national differences) are issuing ethical

¹⁰ It would also be possible to connect data from the HSD to data from two other databases, one with genealogical data that have been processed from public genealogical records and one database of genetic information that has been processed from biological samples obtained for research by physicians cooperating with deCODE.

¹¹ V. Arnason, coding and consent: moral challenges of the database project on Iceland, 2004 and David E. Winickoff, *Genome and Nation Iceland's Health Sector Database and its Legacy*, 2006: http://ecnr.berkeley.edu/vfs/Pls/Winickoff-DE/web/GENOME_NATION.pdf

¹² http://www.norden.org/en/publications/publikationer/2008-570/at_download/publicationfile

recommendations on government policies, regulatory needs to governments and various public institutions. Another objective is to “educate the public” and stimulate the public debate by for instance providing input to journalists, and by organising public meetings about relevant topics.¹³ The Nordic Committee on Bioethics is a very good example of an existing joint Nordic cooperation in this respect, but there is a need for a joint Nordic ethical perspective also in other research areas if Nordic collaboration on registers and public databases is to become a reality. In a wider Nordic database and register collaboration, the national ethics advisory boards could be given the task to lift the ethical debate and societal trust to a Nordic level, as described in the previous section.

The role of Research Ethics Committees

The research ethics committees (the name differs slightly in the Nordic countries) preview research protocols and oversees the compliance by specific research projects according to national legislation. Among many issues it covers the approval of research that involves dealing with sensitive information about people or personal information. Until recently, research ethics committees have been mainly (bio)medical research ethics committees. Only a few ethics committees have concentrated e.g. sociological and health research.¹⁴ For a reinforced Nordic cooperation on registers and databases there is a need to decide on the roles and responsibilities of the ethics committees for research conducted on a Nordic level, and to align their working methods. This is closely linked to harmonization of national legislation on the Nordic level, described in the chapter dealing with the legal perspective.

4.2.5 The technical perspective

The technical challenges for a Nordic reinforced data resources depends largely on the decided level of collaboration, but it includes some sort of joint infrastructure or interoperability mechanisms between the respective national data infrastructures. The key from a technical perspective is precisely the interoperability. Collaboration could be made possible via for instance semantic interoperability on metadata level and interoperable services, like the suggested metadata catalogue mentioned in the organizational perspective chapter on page 17.

There are very few proposed technical solutions which would support Nordic collaboration on data resources in general, but one report

¹³ The descriptive information on ethical advisory boards is based on the book *Bioethics, Politics and Business*

¹⁴ The descriptive information on the research ethic committees is based on the book *Bioethics, Politics and Business*

has currently been released which partly handles the issue: “Research in the Nordic countries – a proposal for increased collaboration, a report from NordForsk’s Task Force for Access to national data repositories.”¹⁵ The Task Force states that Nordic national or joint infrastructures for safe database handling, retrieval and analysis of data do not yet exist. The Task Force also touches upon possible technical solutions for a common system of shared government and researcher owned data for research purposes. One of the recommendations is for NordForsk to set up a Nordic Center of Excellence for the purpose of collaborative research and data sharing, called Nordic Center for Data sharing in Health research (NCDH). Initially, it should focus on collaborating within the health sciences, but sister organizations may be created later within for instance the social sciences, climate and environment. The report suggests Nordic federated data rather than large merged databases, since it would involve combining data from large and sensitive registers that cannot be physically moved. A federated database system is a type of meta-database management system which integrates multiple autonomous database systems into a single federated database. 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 a distributed system, the data structure needs to be reasonably compatible for the researchers to 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 in a way that makes that it is possible to link data across organizations and data sources. The report also suggests to explore the possibility to use cloud computing in association with distributed data sources.

Technical solutions need political support

The main challenge from a technical perspective is the lack of clear political decisions on the level of Nordic collaboration. The technical development is moving very fast and will most likely be able to find solutions to the needs, but there is no possibility to develop large scale technical solutions adapted to the Nordic conditions unless there is a clear long-term political support which clearly states the development of specific technical solutions. Technical pilot test are very useful, such as the proposed establishment of the “Nordic Center for Data sharing in Health research” but the political ambition of turning it into a sustainable, solution which will be scalable has to be clear from the beginning. Otherwise

¹⁵ http://ki.se/content/1/c6/13/68/04/Sharing_Registry_Data_for_Health_Research_in_the_Nordic_Countries.pdf

the risk is that yet another joint infrastructure is built up but not maintained after the project period or test period.

Possible technical solutions in EU projects

An ongoing project where inspiration to some of the technical challenges may be found is the “Data without Boundaries (DwB)”, which is explained more in detail in the last chapter. The project supports equal and easy access to official microdata for the European Research Area and will propose a pilot system for a European distributed remote access to confidential microdata for national datasets. DwB also takes part in the discussion about metadata standards with the aim of building a single point of access for official microdata.

The technical solutions in the EUDAT project, described in the last chapter, also need to be monitored carefully. Since the project is coordinated by CSC – IT Center for Science, Finland and involves Norwegian and Swedish partners, there is a natural link to the Nordic collaboration. EUDAT aims to develop a collaborative, generic infrastructure for scientific data management that can be used by a diversity of research communities and existing infrastructures. Although the project does not target the challenges with registers, the project will try to combine 25 European partners, including data centers, technology providers, research communities and funding agencies from 13 countries for a Collaborative Data Infrastructures, which is a major technical challenge.

4.3 Summary of challenges

The perspectives described are shortly summarised below:

The organisational perspective

- Scattered data resources
- Authorities lacking research perspective
- The lack of a common language – researchers versus authorities
- Specific challenges for register-based research – access to personal data
- Researchers hesitation towards data sharing despite existing infrastructures
- Lack of Nordic perspective in national strategies targeting research data resources

The legal perspective

- Different legislation regarding mainly personal data in different Nordic countries and between different authorities
- The revision of the EU Data protection legislation – how will it affect the national legislation and possibilities for register-based research?

The financial perspective

- The importance of cost-efficient research
- Not all existing data resources can survive
- Funding incentives for data sharing
- Pooling of national funds for Nordic research benefit

The ethical perspective (mainly regarding register-based research and biobanks)

- Keeping the societal trust
- Create an ongoing open ethical debate on the Nordic level
- Decide on the role of ethical advisory boards and ethics committees

The technical perspective

- How to go from pilot project to large-scale infrastructure solutions
- Technical solutions need political support
- Importance of monitoring technical solutions in other initiatives and projects

Estimated challenge level

Level of challenge	Political	Organisational	Legal	Financial	Ethical	Technical
High	x	x	x			
Medium				x	x	x
Minor						

The table divides the different challenges into a high, medium or minor level. None of the challenges are marked as minor, but the Ethical, Financial and Technical challenges are valued as medium challenges compared to the Political, Legal and Organisational challenges.

The Political challenges are valued as high, since this is considered to be the basic requirement for all the other challenges. Unless there is a clear political mandate for a reinforced Nordic collaboration, the other challenges will remain unsolved. The political mandate should be as clear as possible and state the actual meaning of “reinforced Nordic collaboration” on data resources. A clear political statement makes it easier to solve the financial challenges and to formulate necessary policies and regulations which will tackle the organisational and the legal challenges.

The organisational and legal framework are both substantial and usually require a long time to be implemented and a long time to get used to.

The financial challenges and budget allocations depend largely on the political need and support, but where there is a will there are usually funds.

The ethical challenges are marked as medium, since there is a coherent Nordic ethical view in general among the public and among the decision-makers regarding sensitive ethical issues like biobanking and sharing of personal data for research purposes.

The technical challenges are also marked as medium. Sustainable data infrastructures, data sharing, re-use and data storage are topics which have gained much interest during the last few years, with many interesting initiatives including new technical solutions both inside and outside the Nordic countries. The technical challenges can probably be tackled as long as there is a clear vision on the aim of a Nordic reinforced collaboration on databases and registers.

5. National progress

The following chapter describes some of the main progress in sharing data resources useful for research and developing register-based research in all the five Nordic countries. It is not a comprehensive description but an overview of some of the main developments in the field.¹⁶ The general key players from a policy perspective on the national level regarding particularly databases and register-based research are summarised in the table on the following page.

The national overview shows significant progress on the national level in all the five countries and increased awareness of these national initiatives can be useful in a Nordic perspective, if concrete actions are taken to actively learn from best practices and incorporate the results in future Nordic work on shared data resources.

¹⁶ Part of the information describing the development before 2010 is based on an Interim report by the Nordic Research Infrastructure Network NRIN, funded by NordForsk
<http://www.nordforsk.org/files/nrin-interim-report-december-2010>

	Denmark	Finland	Iceland	Norway	Sweden
Research infrastructure on ministerial level	Ministry of Science, Innovation and Higher education	Ministry of Education and Culture	Ministry of Education, Science and Culture	Ministry of Education and Research	Ministry of Education and Research
Research infrastructure policy and funding	Danish Agency for Science, Technology and Innovation (DASTI)	Academy of Finland	Icelandic Science and Technology Policy Council,	The Research Council of Norway	Swedish Research Council,
Bodies			The Icelandic Centre for Research (Rannis)		DISC
Research data service providers	The Danish Data Archive (DDA)	CSC –IT Center for Science, Finnish Social Science Data Archive (FSD)		Norwegian Social Science Data Services (NSD)	Swedish National Data Service (SND) And ECDS (environmental data)
National statistics institutes	Statistics Denmark (remote access to micro data)	Statistics Finland	Statistics Iceland	Statistics Norway	Statistics Sweden (MONA: remote access to micro data) (The National Board of health and welfare)
National data protection commissioners	Danish Data Protection Agency	Data Protection Ombudsman	Icelandic Data Protection Agency	The Data Inspectorate	The Swedish Data Inspection Board
national research ethics	The Danish council of ethics, The National Committee on Health Research Ethics	National advisory board on research ethics in Finland, various ethical delegations	Iceland plans to set up National Advisory Board on Research ethics	National research ethics committees,	The Central Ethical Review board,
Bodies			National bioethics committee	The Norwegian Biotechnology Advisory Board	Regional ethical review boards to approve specific research projects

5.1 Denmark

The Danish Ministry of Science, Innovation and Higher education has the overall responsibility for Danish Research Infrastructure and supervises the research policy and funding body *Danish Agency for Science, Technology and Innovation (DASTI)*.

Statistics Denmark provides researchers with remote access to micro data. Only Danish research environments are granted authorisation of access to micro data. To reduce the cost of datasets for research purposes and solve special data problems, Statistics Denmark has set up a number of research databases, such as the Integrated Database for Labour Market Research (IDA), the Demographic Database, the Fertility Database, the Prevention Register (health data), and the Social Research Register.

A research project involving processing of sensitive personal data has to be approved by the *Danish Data Protection Agency* before such processing can start.

The Danish Data Archive (DDA) gathers, preserve and disseminate mainly quantitative data created by researchers from social science. Through an agreement with the Danish Data Protection Agency DDA preserves data materials containing personal identifiers. Data and the personal identifiers are stored separately and a special permit is required for access to the data.

5.1.1 Danish Strategies and initiatives

A *Danish Research Infrastructure Roadmap*¹⁷ was produced in April 2011. One of the new aspects with the Roadmap is the mid- and long-term perspective. Previously, short funding periods meant that long term operational funding had to come from the institutions themselves. RI projects regarding databases typically have low investment costs and high operational costs, which made these projects especially hard to fund. A longer time perspective in the new Roadmap has addressed this issue. eScience and register-based research is highlighted in the Roadmap and both a reorganisation and a reinforcement on register-based research will be prioritised by establishing a *Coordinating Organ for register-based research (KOR)*.

Ministry of Science, Technology and Innovation has appointed an *Open Access Committee* and placed this under the steering committee for

¹⁷ <http://en.fi.dk/publications/2011/danish-roadmap-for-research-infrastructure-2011/uk-roadmap.pdf>

Denmark's Electronic Research Library (DEFF). The final report¹⁸ from the Open Access Committee recommends national planning of free access to as well as long-term preservation of primary research data by establishing an inter-disciplinary cooperation organisation that gathers together stakeholders within the field primary research data, such as universities, Danish Data Archive, national libraries, research libraries and researcher groups. It is also suggested to look at similar activities via NordForsk and EU initiatives.

Action plan for accessible research data in Denmark, Germany, the Netherlands and the UK

Another recent report in the field is "A Surfboard for Riding the Wave – Towards a four country action programme on research data".¹⁹ It is an Action plan on making research data accessible in Denmark, Germany, the Netherlands and the UK. The report is a response to the Riding the Wave report²⁰ which was published by the High Level Expert Group on Scientific Data. The report focuses on four key drivers: incentives, training, infrastructure and funding of the infrastructure in relation to further developments in data logistics. The Danish partner involved is Denmark's Electronic Research Library (DEFF).

5.2 Finland

The *Ministry of Education and Culture* has the overall responsibility for the Research infrastructure in Finland. *Academy of Finland* is a governmental funding body for scientific research, including research infrastructure in Finland. The Ministry administers *CSC – IT Center for Science*. CSC is a non-profit company which provides IT support and resources for academia, research institutes and companies. This centralized coordination does not exist in the other Nordic countries. CSC provides access to over 70 various scientific databases in fields such as biosciences, language research, Nanoscience and physics.

The Finnish Social Science Data Archive (FSD) is a national resource centre for social science research and teaching. FSD archives, promotes and disseminates digital research data and is funded by the Ministry of Education. Neither The Finnish Social Science Data Archive nor *Statistics Finland* has the same possibilities to give the researchers access to register data as their counterparts in for instance Sweden and Denmark. This

¹⁸http://www.bibliotekogmedier.dk/fileadmin/publikationer/publikationer_engelske/open_access_2010/pdf/Open_Acces_UK.pdf

¹⁹ <http://www.knowledge-exchange.info/Default.aspx?ID=469>

²⁰ <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

is mostly due to legal restrictions or interpretations. Authorisation is applied for separately from each register controller according to the register's procedures. Measures are currently taken to amend the Finnish legislation, to make it more in line with for EC regulation regarding directions to national statistical institutes. The regulation states that researchers' access to confidential data may be granted for scientific purposes. These measures are described more in the section below.

5.2.1 Finnish strategies and initiatives

Finland produced a *Research Infrastructure Roadmap*²¹ in 2009 and an update is planned by Academy of Finland in 2012. One recommendation in the Roadmap regarding data is that "the usability of national registers and the availability of materials should be improved and costs to the user should be reduced, where necessary by amending related legislation. Valuable bodies of material collected in Finland should be made available for broader international use by increased digitization of materials and by implementing uniform collection procedures in accordance with international standards."

Finland established a *Task Force for utilization of electronic data* in 2011, which has produced a roadmap for the utilization of electronic data in research. Some proposed actions from this roadmap are:

- drafting of a national data policy, which includes international scientific cooperation
- to promote the widest possible degree of openness, availability and usability of data with several changes in the legislation, such as amendments to statistics act, copyright, payment and pricing of data output
- to express a collective will in the government platform and establish a cross-sector coordination group to enhance the data issues from the research perspective, starting the planning of the information infrastructure for research and the legislative reforms that enhance the wider utilization of data resources

The Ministry of Education and culture is currently formally setting this cross-sector coordination group.

A working group set by the Ministry of traffic and communication, got through a governmental resolution in 2011 on *open governmental data as a principle*, to promote the re-use of public digital material gathered for administrative purposes – not just for researchers but for the general

²¹ <http://www.minedu.fi/export/sites/default/OPM/julkaisut/2009/liitteet/opm02.pdf?lang=en>

public. A working group set by the Ministry of Finance is just giving out a report that shall open the all the data possible (not containing personal data) to everybody.

Finnish Information Centre for Register Research (ReTki) tries to address the difficulties for researchers to access register data, by promoting the use of national registers for research purposes, particularly in social and health sciences and by offering information on registers and trainings on the use of registers in research.

The Midras project – Micro Data Remote Access System took place in June 2009 to April 2011 and was a joint project between ReTki and CSC, funded by the Ministry of Education and Culture. The aim was to find out how register-based research and the availability of register data can be made more effective with the help of remote access and how to improve the availability and linking of confidential sensitive data sets from several authorities. The plan is to jointly develop the services with Statistics Finland, CSC and ReTki during 2012 and 2013 but final decisions still needs to be taken.

5.3 Iceland

The Ministry of Education, Science and Culture has the overall responsibility for the research infrastructure on Iceland and supervises the *Council for Science and Technology* which is the main body responsible for developing and adopting the general policy for science and research in Iceland. The Icelandic Centre for Research (Rannis) cooperates closely with the Council for Science and reports to the Ministry.

Statistics Iceland may grant researchers with access to registers or databanks with sensitive information if the researcher has been granted permission by the *Data Protection Authority* and (if applicable) the *National Bioethics committee*. There is however no remote access to micro-data as in Denmark or Sweden.

5.3.1 Icelandic strategies and initiatives

Iceland produced a *National Research Infrastructure Roadmap* in spring 2009 and is planning an update in 2012. The Roadmap recommends to ensure future national access to databases and electronic journals and to form a task group to lead the future arrangement of databases in Iceland. This initiative was developed more in the *Science and Technology Policy for Iceland 2010–2012*.²² The Task group for future arrangement of databases in Iceland is going to organise the hundreds of diverse scientific databases which exist in Iceland. The main activities are to organise the structure and coordination of the databases, improve the access to data as well as organising the administration and maintenance of databases.

5.4 Norway

The *Ministry of Education and Research* has the overall responsibility for research infrastructure in Norway. *The Research Council of Norway* provides funding for national infrastructure for research under the National Financing Initiative for Research Infrastructure, which had its first call in 2009. The council is also responsible for updating and developing the Norwegian Roadmap for Research Infrastructure and collaborate internationally.

Norwegian Social Science Data Services (NSD) is a Limited Company owned by the Ministry of Education and Research. The Research Council of Norway has provided the resources required for developing NSD as a national infrastructure facility for access to data for research. Together with *Statistics Norway*, NSD either own or administer most of the registry-based administrative data as well as other large databases on social conditions in Norway. Statistics Norway has chosen NSD as a channel for data dissemination for research purposes, which is a unique solution in an international perspective. NSD is also the Data Protection Official for research and has signed agreements with all Norwegian universities and a number of health enterprises and research institutes. This means that *the Norwegian Data Inspectorate* has delegated responsibility to NSD in relation to the Personal Data Act and Personal Health Data Filing System Act.

²² <http://eng.menntamalaraduneyti.is/media/MRN-PDF-Althjodlegt/STPC-2010-12-Final.pdf>

5.4.1 Norwegian strategies and initiatives

The latest *Norwegian Roadmap for Research Infrastructure* was updated in February 2012.²³ It is closely linked to the *National Financing Initiative for Research*, which states that “Scientific databases and collections” is one of the categories eligible for funding within the field Research Infrastructure. In the RI Roadmap it is stated that the Research Council of Norway will contribute to the work of making research data available in safe systems and in such format that they can form the basis for research, both nationally and internationally, as well as ensure Norwegian participation in international computer networks.

The Research Council granted NOK 35 million for developing the *Remote Access Infrastructure for Register Data (RAIRD)* in October 2011. The new infrastructure is designed to support social science research, but it will also be suitable for other research fields that utilise register data, such as environmental science, cultural studies, and medical and health research. In Norway, the current procedures for access to registry data are time-consuming and resource-intensive, mainly because the data involved are highly sensitive, and there are different registry owners. RAIRD will address the aspects of technology, organisation, administration and dissemination of research data. The objective is also to make it easier for international researchers to make use of Norway’s register data. NSD and Statistics Norway will develop the RAIRD infrastructure together and the new system will be available to all researchers and research institutions.

The Research Council of Norway is also launching a new research programme “*Biobanks and health data*”. The aim of the programme will be to use full research potential of biobanks by linking the results of analyses from the biobanks to data collected from health surveys, health registries and health services. The programme seeks in particular interdisciplinary projects. The first funding announcement is planned for early 2012 and the first programme period is 2012–2016.

²³<http://www.forskningradet.no/servlet/Satellite?c=Nyhet&pagename=infrastruktur%2FHovedsidemal&cid=1253973200279&p=1224697900480>

5.5 Sweden

The *Ministry of Education and Research* has the overall responsibility for research infrastructure in Sweden. The *Swedish Research Council* is involved in financing and long-term planning for national research infrastructure as well as Sweden's participation in international research infrastructures.

The Swedish Research Council has formed a *Database Infrastructure Committee (DISC)* with the task of co-ordinating existing and new, quality-assured research databases and to work on matters to do with accessibility, quality assurance and integrity. DISC has appointed the *Swedish National Data Service (SND)* as a national resource for the coordination of existing and newly established databases within the social sciences, humanities and health sciences. SND offers support to Swedish research by facilitating researchers access to data and to support archiving, usage and re-usage of databases and related research materials. SND is still building up its activities and one important mission is to gain a good overview of existing research databases and to gain full knowledge of future databases within the social sciences, humanities and social medicine.

Statistics Sweden provides researchers with register data and remote access to microdata via the Microdata Online Access (MONA). The opportunity to communicate data via the Internet to other countries is limited due to that the Swedish law cannot protect data transmitted to another country. When sensitive personal data should be handled for research purposes, data delivery must be approved by an ethics board. The general rule is that Statistics Sweden only discloses anonymised data. *The National Board of Health and Welfare* is a government agency under the Ministry of Health and Social Affairs. Even though statistics is not the main task, the National Board of Health and Welfare publishes statistics in the areas of health, medical care and social services and maintains health data registers, which can be used by researchers. Similar rules apply as when applying for research data from Statistics Sweden.

5.5.1 Swedish strategies and initiatives

Sweden has also produced a national *Research Infrastructure Roadmap* from 2007 but a more current strategy document is the latest *Swedish Research Councils guide to research infrastructure*²⁴ published in the beginning of 2012. The document contains suggestions to the Swedish government on future priorities of research infrastructure for the next

²⁴<http://www.vr.se/download/18.9b19d771342fcd4fc8000817/Vetenskapsradets+guide+till+infrastrukturer+2012.pdf>

research and innovation government bill. Some of the actions to develop Swedish research are related to databases and registers:

- Sweden should take advantage of its position in research involving personal registers for studies of urgent issues regarding the correlation between social conditions, economics, health and education, as well as the molecular and lifestyle mechanisms behind common diseases in our population
- Sweden should be a driving force behind the co-ordination of infrastructure and data in the area of climate and the environment
- Sweden should develop and co-ordinate biomedical infrastructures by developing existing national infrastructures, constructing new ones, and promoting co-ordination and data integration between them

The Nordic globalisation initiative and the collaboration regarding specifically registers and biobanks are also mentioned. A detailed report from DISC and the Swedish Research Council on legislative preconditions for a Swedish database infrastructure (2010) suggested *a review in legislation to support register based research*, which is currently going on. *SIMSAM – Swedish Initiative for research on Microdata in the Social and Medical Sciences* is an initiative supported by the Swedish Research Council. SIMSAM promotes improved and expanded use of registers in research and aims to increase the knowledge of how registry research is organized in Sweden and internationally. SIMSAM prioritises interdisciplinary, innovative register-based research and six “research nodes” and one graduate school have been granted funding for five years from december 2008 until november 2013.

6. International view

The European Strategy Forum, ESFRI is the member states platform for planning and coordination of joint European infrastructure. ESFRI's "European Roadmap for Research Infrastructures" sets the course for joint European Infrastructure within a range of science fields. The latest Roadmap was produced in May 2011 and most countries (including the Nordic countries as shown in the previous chapter) have produced their own national Research Infrastructure Roadmaps in parallel to the ESFRI Roadmap.

The Nordic countries are involved in many of the ESFRI infrastructure projects and in the Nordic Research Infrastructure Network (NRIN) Interim report from December 2010,²⁵ the ESFRI projects have been scanned in order to identify those projects worth pursuing at Nordic level. Although the list is not a total scan it gives a good indication on where at least three of the Nordic countries have expressed an interest in the ESFRI project, which is mainly in the area of environment and health. The criteria for the scanning were also that the ESFRI project must seem to have a potential for further Nordic collaboration and bring an added value. Some of the suggested ESFRI projects involving data infrastructure collaboration are *CESSDA* – the umbrella organisation for social science data archives across Europe, the biobank infrastructure *BBMRI.eu* (where the Nordic perspective already exists), the bioinformatics infrastructure *ELIXIR* and the Biodiversity infrastructure *Lifewatch* – where NordForsk has recently decided to fund a Nordic network similar to the BBMRI – Nordic.²⁶ These are projects where the Nordic cooperation possibly could be developed as a "subcooperation" within the bigger project to adopt the solutions to the Nordic collaboration and to create a Nordic joint voice in a European context.

There are also other projects, where only a few of the Nordic countries are involved, but where best practices and experiences may be transferred to the Nordic collaboration:

²⁵ <http://www.nordforsk.org/files/nrin-interim-report-december-2010>

²⁶ More information will soon be available on NordForsk website: <http://www.nordforsk.org>

GRDI2020 – Towards a 10-Year Vision for Global Research Data Infrastructures

Duration: 1 February 2010 – 24 months
Nordic Partner: The Royal Institute of Technology, Sweden

Objectives

Contributing to the definition and development of a European policy for sustainable Global Research Data Infrastructures. Support data movement across scientific disciplines to create interdisciplinary science, with the aim to reduce duplication of effort and resources. GRDI2020 has recently produced a 10 year visionary roadmap for Global Research Data Infrastructures, which outlines the framework for delivering Data Infrastructures that meet the needs of the Research Community.

- Website: www.grdi2020.eu

EUDAT

Duration: 1 October 2011 – 36 months
Nordic Partner: CSC – IT Center for Science, Finland (coordinator), SIGMA, Norway, SNIC, Sweden

Objectives

EUDAT aims to develop a collaborative infrastructure “a pan-European data eco system” for scientific data management that can be used by a diversity of research communities and existing infrastructures. EUDAT aims to shape a platform for shared services that enables data-intensive research potentially crossing all disciplines of science. EUDAT will also ensure that the data infrastructure is robust enough to handle the expected acceleration of generated scientific data. The project involves 25 European partners, including data centers, technology providers, research communities and funding agencies from 13 countries.

- Website: www.eudat.eu

Opportunities for Data Exchange (ODE)

Duration: November 2010 – 24 months

Nordic Partner: CSC – IT Center for Science, Finland

Objectives

ODE aims to “add an interoperable data sharing, re-use and preservation layer to the emerging eco-system of e-Infrastructures” to fully enable eScience. The objective is to promote best practices in sharing, re-using, preserving and citing data, by interviewing key players working with policy making, in data centres, data repositories and funding agencies. The collection of the interviews; “ten tales of drivers and barriers in data sharing” can be found on the website. The project is coordinated by CERN.

- Website: <http://www.ode-project.eu/>

Data without Boundaries (DwB)

Duration: 2011–2015

Nordic Partner: University of Tampere, Finland Norwegian Social Science Data Services (NSD), University of Gothenburg (Sweden), Statistics Sweden

Objectives

The project supports equal and easy access to official microdata for the European Research Area. DwB contributes to the creating of an integrated model where the best solutions for access to official statistical microdata are available, irrespective of national boundaries, yet flexible enough to fit national arrangements. Currently, comparative research projects based on national official microdata need to go through multiple accreditations and quite different technical and methodological environments. DwBs aim is to propose pilots for a European research accreditation and a European distributed remote access to confidential microdata for national datasets. DwB also takes part in the discussion about metadata standards with the aim of building a single point of access for official microdata.

- Website: <http://www.dwbproject.org/>

e-Infrastructure Reflection Group (e-IRG)

Important information about data infrastructures can also be found in material from the e-Infrastructure reflection group, where Sweden, Denmark, Norway and Finland are represented. The main objective of the e-Infrastructure initiative is to support the creation of a political, technological and administrative framework for an easy and cost-effective shared use of distributed electronic resources across Europe. It consists of official government delegates from all the EU countries. The e-IRG produces white papers, roadmaps and recommendations. The e-IRG White Paper from 2011 contains a chapter on data infrastructures: <http://www.e-irg.eu/>

7. References

- Bioethics, Politics and Business, Edited by Salla Lötjönen and Helena von Troil. *TemaNord* 2008:570. Nordic Council of Ministers. http://www.norden.org/en/publications/publikationer/2008-570/at_download/publicationfile
- Coding and consent: moral challenges of the database project on Iceland. V. Arnason 2004
- Dansk roadmap for forskningsinfrastruktur 2011. Forsknings- og Innovationsstyrelsen, april 2011. <http://www.fi.dk/publikationer/2011/dansk-roadmap-for-forskningsinfrastruktur-2011>
- Eurobarometer 73.1 Europeans and Biotechnology in 2010. Winds of change? A report to the European Commission's Directorate-General for Research, October 2010. http://ec.europa.eu/public_opinion/archives/ebs/ebs_341_winds_en.pdf
- Genome and Nation Iceland's Health Sector Database and its Legacy, 2006. David E. Winickoff. http://ecnr.berkeley.edu/vfs/PIs/Winickoff-DE/web/GENOME_NATION.pdf
- Knowledge Infrastructure for the Fifth Freedom in the Baltic Sea Area. By Erik Elmroth (scientific Secretary). Nordic Council of Ministers, Copenhagen 2010. <http://www.norden.org/en/publications/publikationer/2010-516>
- National-level research infrastructures, present state and Roadmap. Finnish Ministry of Education 2009. <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2009/liitteet/opm02.pdf?lang=en>
- Nordic eScience Action Plan. 10 concrete actions to implement the Nordic eScience Strategy. Nordic eScience Initiative (eNORIA), 2008. <http://www.nordforsk.org/en/publikasjoner/nordic-esience-action-plan/view>
- Nordic eScience conference 29–30 September 2011 arranged as a part of the Finnish Presidency in the Nordic Council of Ministers. <http://www.esceince.fi>
- Nordic eScience Research, Education, and Sustainable Infrastructure Services. A strategy document for the Nordic Council of Ministers 2007–07–17. http://www.vr.se/download/18.6f6ddf01189d73e59380002244/nordic_escience.pdf
- NORIA-nettet NRIN, Nordic Research Infrastructure Network Interim report 1 Project year, 2010. <http://www.nordforsk.org/files/nrin-interim-report-december-2010>
- Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf
- Recommendations for implementation of Open Access in Denmark. Report from the Open Access Committee. Published in 2010 by Denmark's Electronic Research Library, Danish Agency for Libraries and Media http://www.bibliotekogmedier.dk/fileadmin/publikationer/publikationer_engelske/open_access_2010/pdf/Open_Acces_UK.pdf
- Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission October 2010 <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- Rättsliga förutsättningar för en databasinfrastruktur för forskning. Vetenskapsrådets rapportserie 11:2010 <http://www.vr.se/download/18.28ca860b1293e247fc480006831/Report+11+2010.pdf>
- Science and Technology Policy for Iceland 2010–2012. Prime Minister's Office, Science and Technology Policy Council, 2010. <http://eng.menntamalaraduneyti.is/media/MRN-PDF-Althjodlegt/STPC-2010-12-Final.pdf>

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 by Magnus Stenbeck, Ulf Jonsson, Sweden, Maria Heimisdottir, Iceland, Mads Melbye, Denmark, Marjut Salokannel, Finland, Camilla Stoltenberg, Norway. http://ki.se/content/1/c6/13/68/04/Sharing_Registry_Data_for_Health_Research_in_the_Nordic_Countries.pdf

A Surfboard for Riding the Wave. Towards a four country action programme on research data. Knowledge Exchange, Maurits van der Graaf, Leo Waaijers, November 2011. <http://www.knowledge-exchange.info/Default.aspx?ID=469>

Survey on open access in FP7. Directorate-General for Research and Innovation, 2012. http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-open-access-in-fp7_en.pdf

Verktøy for forskning – del I och del II. Nasjonal strategi for forskningsinfrastruktur 2012–2017 <http://www.forskningsradet.no/servlet/Satellite?c=Nyhet&pagename=infrastruktur%2FHovedsidemal&cid=1253973200279&p=1224697900480>

Vetenskapsrådets guide till infrastrukturen 2012. Rekommendationer för forskningsinfrastruktur på lång sikt från forskningsråden och Vinnova. Vetenskapsrådet 2012. <http://www.vr.se/download/18.9b19d771342fcdf4fc8000817/Vetenskapsradets+guide+till+infrastrukturen+2012.pdf>

8. Sammanfattning

Denna studie kartlägger vilka utmaningar som finns för ett förstärkt nordiskt samarbete gällande dataresurser. Med dataresurser menas främst befintliga nationella databaser och register upprättade huvudsakligen för administrativa ändamål, men även ny forskningsdata behandlas. Utmaningarna analyseras från ett politiskt, juridiskt, etiskt, organisatoriskt, tekniskt och ekonomiskt perspektiv. Studiens breda fokus riktar sig främst till beslutsfattare som arbetar med e-vetenskap och e-infrastruktur på nationell och/eller nordisk nivå. Involverade parter är Nordiska ministerrådet, NordForsk, finska Undervisnings- och kulturministeriet samt CSC -IT-centret för vetenskap.

Särskilt fokus har lagts på personnummerbaserad registerforskning, eftersom alla de nordiska länderna har en stor mängd myndighetsregister som ännu inte används till sin fulla potential inom varierande forskningsområden. En stor utmaning är att skapa möjligheter att kombinera dessa register med annan forskningsdata och de nordiska länderna måste i detta fall utveckla sina egna unika samarbetsmetoder eftersom det inte finns några likvärdiga förutsättningar för denna typ av forskning i resten av Europa.

De största utmaningarna från ett *organisatoriskt* perspektiv är de utspridda dataresurserna, vilket både syftar till den stora mängden databaser och register samt de många skilda organisationer som arbetar med att samla in och upprätthålla dessa dataresurser. Andra organisatoriska utmaningar är bristen på ett naturligt forskningsperspektiv hos berörda myndigheter samt avsaknaden av ett gemensamt språk mellan forskare och myndighetsanställda. För registerbaserad forskning är det dessutom ofta en komplicerad och utdragen process att få tillgång till känsliga personuppgifter. Att dela med sig av och återanvända ny forskningsdata kan också vara komplicerat, delvis eftersom forskarna själva ofta tvekar att dela med sig av sin insamlade data, trots att de har tillgång till nödvändig datainfrastruktur. Bristen på ett uttalat nordiskt perspektiv i nationella forskningsstrategier utgör också ett hinder för ett förstärkt nordiskt samarbete, eftersom det finns en risk att utveckla politiska direktiv och riktlinjer som inte tar hänsyn till de nordiska förutsättningarna.

Från ett *rättsligt* perspektiv ligger de största utmaningarna i de nordiska ländernas skiftande lagstiftning, främst gällande forskarnas möjligheter att få tillgång till persondata. Lagstiftning och praxis skiftar både mellan länderna samt mellan olika myndigheter i samma land. Den pågående översynen av EU:s regler för uppgiftsskydd och hur det kommer att påverka den nationella lagstiftningen och möjligheter till att utöva

registerbaserad forskning skapar också en viss osäkerhet, även om en ny EU-förordning med krav på översyn av de nationella dataskyddslagarna också kan vara ett utmärkt tillfälle för de nordiska länderna att se över och harmonisera lagstiftningen ytterligare på nordisk nivå.

Det *ekonomiska perspektivet* belyser fördelarna med kostnadseffektiv forskning. En stor mängd offentliga medel används för att upprätthålla befintliga offentliga register och databaser samt att samla in ny data och dessa dataresurser bör därför användas till sin fulla potential inom forskningen.

Det bör dock utvärderas vilka dataresurser som är värda att anpassa och göra långsiktigt hållbara som en del av en nordisk datainfrastruktur. Ekonomiska incitament kan också användas för att uppmuntra forskare att dela med sig av sin forskningsdata. Från ett ekonomiskt perspektiv är det också viktigt med en sammanslagning av nationella medel för att långsiktigt kunna utveckla det nordiska samarbetet. Samtidigt finns det många andra internationella initiativ som kräver stora ekonomiska satsningar och en lösning skulle kunna vara att delta i exempelvis ESFRI projekt och samtidigt främja ett nordiskt perspektiv inom dessa EU-initiativ genom att inkludera ett "nordiskt samarbetslager" inom det större EU samarbetet.

Det *etiska perspektivet* hanterar främst registerbaserad forskning och biobanker där utmaningen är att kunna ta tillvara de unika nordiska förhållandena för registerbaserad forskning utan att inkräkta på den personliga integriteten. En viktig aspekt är att befolkningen har ett fortsatt högt förtroende för denna typ av forskning och att det förs en kontinuerlig, öppen etisk debatt, inte bara på nationell nivå utan även på nordisk nivå genom att exempelvis skapa fler nordiska kanaler för offentlig dialog och medborgarengagemang. De nationella etiska råden och kommittéernas roller och arbetsmetoder måste också förtydligas och utvecklas vid ett förstärkt nordiskt samarbete.

Det *tekniska perspektivet* behandlar främst utmaningar gällande interoperabilitet och möjligheterna att ta steget från lovande pilotprojekt till storskaliga lösningar. Det tekniska perspektivet betonar också vikten av att ha ett uttalat politiskt stöd för att kunna utveckla tekniska lösningar och möjligheten att bevaka och tillvarata nya tekniska lösningar inom internationella initiativ och projekt.

Studien ger också en nationell överblick av den aktuella utvecklingen i de fem nordiska länderna för att öka kunskapen om viktiga nationella dataresursinitiativ som kan bidra till en starkare samverkan på nordisk nivå. Det sista kapitlet beskriver några pågående intressanta initiativ utanför de nordiska länderna såsom "EUDAT" och "Data Without Boundaries". I några av dessa initiativ deltar de nordiska länderna men som enskilda länder snarare än med ett nordiskt perspektiv.

Reinforced Nordic collaboration on data resources

Challenges from six perspectives

This study is mapping the most significant challenges and obstacles for a reinforced Nordic cooperation on data resources. Focus is put on existing national databases and registers established mainly for administrative purposes but also the question of newly-generated scientific data is handled. The challenges are analyzed from political, legal, ethical, organisational, technical and financial perspectives. The broad scope targets primarily policy makers involved in eScience development on national and/or Nordic level. Involved parties in the study are Nordic Council of Ministers, NordForsk, Finnish Ministry of Education and Culture and CSC – IT Center for Science.

Special focus has been put on the challenges for register-based research, since all the Nordic countries have a vast amount of population-based registers which are considered not to be used to their full potential in various research fields. One of the challenges is to find way to combine these registers with other research data and the Nordic countries have to find their own way of unique collaboration methods, since there are no equivalent pre-conditions for this kind of research in the rest of Europe.

The study also gives a national overview of the current progress in the five Nordic countries, to raise awareness of important national initiatives which can contribute to a stronger collaboration on the Nordic level.