

SHARING RESEARCH DATA AT THE NORDIC LEVEL

- A vision for the future Nordic cooperation from Finnish perspective

Marjut Salokannel

Current situation in Finland

In Finland open government data and data produced by public funding has been on the foreground during the past years. At the governmental level there have been several working groups dealing with these issues, one of which has specifically concentrated on opening-up publicly funded data for research purposes. A continuation of this work is carried on by a project of the Ministry of Education and Culture.¹

At the international level research funders around the world are requiring that the results of the research they have funded have to be publicly accessible. This refers both to the publications and the underlying data. In Finland the national research funding agency, the Academy of Finland, is not there yet, but is merely recommending that research articles be openly available. The Academy does require that the application for funding includes a data management plan. When we look at the general picture with regard to the availability of research data in Finland there are big differences depending on the field of research. In natural sciences and biomedicine open access to research data is the norm and the data is being deposited in the international depositories whereas in humanities and social sciences, with the exclusion of economics, open research data remains still a rarity.

Data sharing at the Nordic level

Even if the general approach with regard to opening up research data follows the internationally established practices of each research field, there are some issues that are specific to the Nordic countries. One of these issues relates to the vast amounts of register data contained in different population based registers which are almost identical in all the Nordic countries. These registers are an immense source of raw data for epidemiological research, population based research,

economics and other fields of research where combining and analysing microdata from different sources forms the basis of research. Similarly datamining different sources of data in bioinformatics and genomics gives us new knowledge of diseases and basis for developing therapies.²

Unfortunately, these data reserves are currently being underutilized, since using the microdata contained in these registers for research purposes poses constantly difficulties, in particular with regard to combining data from different sources. In Finland the most difficult bottleneck has been the data stored within the National Statistics Authority.

In the following we shall discuss the core legal issues in respect to accessing research data, that is the extensive copyright protection and data protection law. Finally we raise the question of how it would be possible to establish, at the Nordic level, a secure federated environment for a seamless use of register data from different sources irrespective of the place of residence of the researcher, and a way to combine this data with other data, such as data in biobanks or patient records.

Copyright issues with regard to using data for research purposes

The Nordic copyright laws have a common history dating from the beginning of 1960's. Since then there has been some minor variations in the national laws but the basic principles remain very similar. One of these basic issues is the extensive protection afforded by the law to all sorts of collections of data resulting from a combination of the EU *sui generis* database right and the Nordic catalogue rule which extends the protection also to those databases that do not fall under the copyright protection or *sui generis* database protection. From researcher's perspective the situation

¹ Exploitation of Public Sector Information" -working group in the Ministry of Traffic and Transport, two working groups in the Ministry of Finance ("Interfaces of public sector basic information resources" and "More efficient use of public sector information resources"), a Ministry of Finance working group for amending the Statistics Act and a cross-sectoral steering group by the Ministry of Education and Culture "Information resources for Research".

² See also "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, available at http://ki.se/content/1/c6/13/68/01/Sharing_Registry_Data_for_Health_Research_in_the_Nordic_Countries.pdf (last accessed 14th June 2012).

is aggravated by the fact that, unlike in most other countries in the world, there exists no provision in the law, which would provide legal security for research uses of copyrighted materials. For example, the Finnish Copyright Act does not contain an exception in the law, which would make it possible to use protected materials, including databases or catalogues, for research purposes. This sets Finnish researchers in an unequal position in relation to their foreign counterparts and makes it difficult for researchers residing in Finland to take part in international collaborations.³

Digital technology has made it possible to use data produced by public funding by the government and research institutions in ways that have not been previously possible. This has even given birth to new fields of scientific research such as bioinformatics and biostatistics, which use data mining and analytics as their principal tool for creating new knowledge out of existing data. In this respect the Nordic countries enjoy an especially advantageous position in form of their over century old population based registries which are built on the same basis in all the Nordic countries.

Data mining and text mining in particular have proven to be problematic in terms of the copyright law. In the United States the *fair use* doctrine makes it possible to use data mining techniques for research purposes but in Europe this is not necessarily the case. In particular in countries, such as Finland, where there is no statutory exception providing for research use of copyrighted material, undertaking data and text mining currently call for careful legal analysis if future legal problems are to be avoided.

Data protection and data sharing

Data mining raises also questions relating to possible data protection issues if it makes it possible to identify persons by combining data from different sources, such as public population based registers. The EU Commission has tackled this issue in its proposal for a general data protection regulation in which it subjects processing of data for research purposes under specific rules.⁴ The proposed regulation makes it possible to create a cross-border framework for data sharing in e.g. register and biomedical research, by requiring the de-identification of individual data when it is used for research purposes. Data which have been de-identified may be used for research purposes in terms of the

³ *Marjut Salokannel*, Using Copyrighted Works for Research Purposes, IPRInfo 2/2009, available at http://www.iprinfo.com/lehti?action=articleDetails&ca_id=662&id=43 (last accessed 14th June 2012).

⁴ 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, COM(2012) 11 final, 25.1.2012.

regulation also without the consent of the data subjects.

The proposed regulation permits processing of personal data for statistical and scientific research purposes

- 1) with the consent of the data subject for one or several specific purposes;
- 2) processing is necessary for the purposes of historical, statistical or scientific research under the following conditions:
 - a) these purposes cannot be otherwise fulfilled by processing data which does not permit or not any longer permit the identification of the data subject; and
 - b) data enabling the attribution of information to an identified or identifiable data subject is kept separately from the other information as long as these purposes can be fulfilled in this manner. (Article 83)

In addition, the regulation provides that when the data processed by a controller do not permit the controller to identify a natural person, the controller shall not be obliged to acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of the Regulation. (Article 10) This provision provides relief for researchers and research institutions acting as data controllers when research is being conducted with de-identified data.

Future vision for data sharing in register based research at the Nordic level

For researchers it is of paramount importance to be able to combine data from different sources, whether it is from different registers, biobanks, or patient records. Researchers also want to pool this data with their colleagues in other countries and use it in collaborative research projects. The best way to protect the privacy of the persons whose information is contained in such data is to anonymize the data. With the advances of technology complete anonymization is becoming increasingly difficult and we shall have to use rough estimates with regard to the probability of identification of data subjects. Moreover, in scientific research there may be exceptional instances when the identification of a certain person could be necessary.

If register data is encrypted in a way that it does not permit identification of data subjects but allows the decrypting of the codifiers in certain cases

(pseudonymisation), this would allow linking of personal register data with data in different registers without compromising the privacy of data subjects. The Finnish data protection ombudsman has proposed instituting a specific personal research identifier for each person through which different data may be combined from different sources. The identifier would replace the social security number. The next step for Nordic purposes would be the creation of Nordic research identification number which could be used throughout the Nordic countries thus abolishing the need for national identifiers.

When envisioning future Nordic cooperation in data sharing we must keep in mind that Denmark and Sweden already have well functioning on-line environments for using register data for research purposes. From researcher's point of view it is, however imperative that data can be combined from different registers with, e.g., molecular data and data in biobanks, and used in collaboration with other researchers across national borders, which is not possible with closed on-line systems. Cross-border data-sharing is, however, precisely what the EU data protection regulation aims to achieve. Sharing encrypted or pseudonymized data for research purposes is an answer to this question. At the Nordic level, this principle could be applied with regard to register data when combined with common anonymization service which could provide the data subjects with specific research identification codes which would replace the national social security numbers.⁵

From the researchers' perspective what is needed is the possibility of mining, analyzing and combining data from different sources, including from different countries, and to be able to exchange and collaborate in international research settings while using this data. We should be able to make available the Nordic register based data for research purposes in such a way that it is possible to link and combine the microdata with microdata in other registers across borders and, ultimately, with data residing in biobanks or even with patient records in such a way that it does not compromise the privacy of individuals whose data is stored in these registers. How this can be achieved in legal, technical and practical terms requires, however, further study.

Marjut Salokannel is LL.D. Docent Helsinki Collegium for Advanced Studies

⁵ See also the report commissioned by the Nordic Council of Ministers *Sandberg, M.*, Reinforced Nordic collaboration on data resources, TemaNord 2012:514.