D2.3: Open Science in the Nordics: Legal Insights

Author(s)	Lars Fischer (NORDunet), Per-Olov Hammargren (SNIC), Francesca Iozzi (SIGMA 2), Pirjo Kontkanen (UH), Troels Rasmussen (DeIC)
Status	Final
Version	1.0
Date	14/08/2020

Document identifier:	
Deliverable lead	University of Helsinki
Related work package	WP2
Author(s)	Lars Fischer (NORDunet), Per-Olov Hammargren (SNIC), Francesca Iozzi (SIGMA 2), Pirjo Kontkanen (UH), Troels Rasmussen (DeIC)
Contributor(s)	
Due date	31/08/2020
Actual submission date	27/08/2020
Reviewed by	Anca Heinola (FMI), Antti Pursula (CSC)
Approved by	
Dissemination level	Public
Website	https://www.eosc-nordic.eu
Call	H2020-INFRAEOSC-2018-3
Project Number	857652
Start date of Project	01/09/2019
Duration	36 months
License	Creative Commons CC-BY 4.0
Keywords	Legal framework, legislation, implementation

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Abstract:

This document will exemplify how legal frameworks may affect the implementation of EOSC standards and policies. Use cases are used to showcase real world impact, by illustrating legal issues encountered in practice by researchers and research groups. The legislation as such is not always the biggest barrier for sharing data but the way the legislation is implemented.

Executive summary

Legal frameworks have been identified as an area that needs to be considered when implementing the EOSC. EOSC aims to facilitate sharing of data and resources across Europe and realise the vision of FAIR data. To realise this, legal frameworks may need to be aligned, and potential legal barriers considered. This deliverable shows that not only the legislation but also institutional policies implemented to ensure legal conformance are a barrier for implementing cross border data sharing.

To illustrate this, the delivery engages with a number of use cases developed in Work Package 5 of EOSC-Nordic. Representatives for the use cases have been interviewed to show and exemplify what kind of

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

legal issues have been recognized and how aforementioned issues may affect sharing data across the borders. This deliverable is as such not a comprehensive legal review of national legislations but compliments other ongoing efforts regarding screening of existing legislation and regulatory issues.

Key findings of the study are:

- Legislation as such is not always a barrier for cross-border data sharing
- Organisations have restrictive policies for cross-border data sharing
- Support for tackling legal issues and processes to handle data from the beginning to enable cross-border data sharing is needed

This deliverable is the first of two deliverables from the EOSC-Nordic projects exploring the legal framework for open science in the Baltics and Nordics. The present deliverable presents analysis of the current state of play. A future deliverable will propose policy recommendations.

The deliverable is part of the *Policies, legal issues and sustainability* effort of the EOSC-Nordic project to foster coordination of and connection between national initiatives at policy level, in the context of EOSC. The deliverable will help policy makers stay informed about issues related to sharing and safe processing of both open and sensitive data across borders. The deliverable will also service as input to the EOSC governance evolution. In particular, the deliverable complement work to assess the European legal framework for EOSC undertaken by the EOSC FAIR Working Group and will be contributed to the WG's work.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Table of Contents

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

1. Purpose and scope of the document

The purpose of this deliverable is to study and exemplify how legal frameworks may affect the implementation of EOSC standards and policies. The implementation plan for EOSC identifies legal frameworks as an area of study in relation to the rules of participation and in the context of governance[1]. Further, within the EOSC interoperability framework legal aspects related to EOSC are touched on[2]. Likewise, the architecture and rules of participation task forces have touched on legal aspects related to EOSC[3].

The use cases adopted in the EOSC-Nordic project are used to investigate in real scenarios how the legal frameworks might impact the EOSC on a Nordic and Baltic regional dimension. Representatives for the use cases have been interviewed in order to understand the legal issues encountered and the impact on the ability to share data across borders for research purposes.

This deliverable is part of EOSC-Nordic effort to foster connection between national initiatives at policy level, in the context of EOSC. The deliverable will inform policy makers about issues related to sharing and safe processing of both open and sensitive data across borders. The findings will be shared with the EOSC-Nordic partners and with policy makers in policy workshops organized by EOSC-Nordic.

The purpose of this document is not to make a comprehensive legal review of national legislations. This shall be done within the frame of studies regarding cross border data flows initiated by the EOSC Secretariat. That study will entail screening of existing legislation in the member states, including an assessment of legal interoperability that will be added in version 2 of the EOSC Legal Interoperability Framework[4], as well as analysis of legal developments that may facilitate or hinder cross-border flow of data.[5] As such this deliverable will complement the effort of the EOSC FAIR Working Group regarding screening of existing legislation and regulatory issues connected to FAIR principles within different EU jurisdictions, and will be contributed to the work of the EOSC FAIR WG.

This document will not provide policy recommendations. Such recommendations will be addressed in a companion deliverable in August 2021.

2. Representation of EOSC-Nordic use cases

All the EOSC-Nordic use cases involve well established communities working in a variety of disciplines, from climate/modelling data to medical data for personalised medicine. Despite the variety of data types, the common denominator among the use cases is the ambition to increase the scientific potential and impact of their research by sharing data over wider regions and domains, possibly across disciplines and communities. The use cases are also used in EOSC-Nordic Work Package 5 to showcase the implementation of the EOSC on the Nordic and Baltic region through a number of demonstrators. Similarly, the use cases serve as real-word scenarios to probe the complexity of the legal and policy landscape while implementing the EOSC, which is activity of EOSC-Nordic Work Package 2 and reported in this deliverable. In both activities (WP2 and WP5) the Baltics and Nordics are adopted as a sandbox to explore potential and possible bottlenecks in the real implementation of the EOSC

---

[1] http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

[2] https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-open-science-cloud-eosc.pdf

[3] https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/ae215698-af7b-11ea-bb7a-01aa75ed71a1

[4] Corcho et.al.: EOSC Interoperability Framework, https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf

[5] https://www.eoscsecretariat.eu/funding-opportunities/open-calls/experts-study-legal-regulatory-issues-fair-principles

The current work is based on information collected through interviews of the Principal Investigator of the use cases. To ensure internal consistency of the collected data, the interviews were conducted following a pre-designed guide. The interview guide (Appendix A) entailed questions about both data and legal issues, solved, and unsolved, encountered by the use cases. The answers were validated with the interviewees to ensure that their answers were properly understood.

The use cases were then used as basis for analysis. Legal issues were identified and their effects on data sharing were studied. Also, other identified aspects related to legal issues were studied.

2.1. Methodology and choice of use cases

To better understand the legal and legal-related issues faced by Nordic and Baltic researchers when engaging in cross-border data sharing, this deliverable takes a qualitative, case-based approach.

A use-case based approach allows an investigation of existing research efforts and the barriers and restriction researchers face when their data-driven research is done in a cross-border setting.

Further, a qualitative study allows us to investigate different aspects of cross-border data sharing and to gain insights into how legal matters impact research. With the qualitative approach, legal implication for different types of data sets, for different ways of sharing the data, and for key research areas can be considered.

The use-case based, qualitative approach considers a limited number of use cases, their data use, and the legal aspects of data sharing. The specific issues in the use cases are studied, and insights are extrapolated for types of data sharing of the use case exemplifies.

To support this, we have selected four use cases. The use cases are selected to cover the following aspects of cross-border data sharing

- different modes of data sharing (a cross-border research collaboration, research accessing an existing data set in another country, and effort gathering a repository of data for later cross-border use).
- Cross-border exploitation of sensitive data as well as data with limited or no privacy concerns
- Cross-border exploitation of medical research and case data, as this is currently a major policy concern of the Nordic countries in the framework of the Nordic Council of Ministers[6]
- Exploitation of data sets with a single owner as well as data sets aggregated from multiple contributors over time
- Research at different types of institutions, such as research hospitals, major universities, and smaller research groups
- Represent a range of Nordic and Baltic countries in the cases

The above mentioned use-cases, even though few, have significant diversity and allow us to consider the impact of key issues such as IPR, access rights, GDPR, national policy for handling sensitive data, and to consider key research areas such as bioinformatics, medicine, artificial intelligence and machine learning, biodiversity, and climate research.

The present study is, as previously mentioned, not meant to be a comprehensive, quantitative study. The legal insights gained are illustrated and their impact discussed. Their frequency or cost is not assessed, and the impact of the various insights is not compared. For this reason, the number of cases is intentionally small to maintain focus on key issues.

---

[6] https://www.nordforsk.org/programs/nordic-programme-health-and-welfare.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Neither is this a study of the full range of legal restrictions or review of policies that may impact cross-border data sharing and research. The legal insights are approached entirely from the perspective of the experience of researchers – not from the perspective of the body of legislation. Such a study in currently under way by the EOSC FAIR working group. These two studies are complementary.

2.2 Case PM Heart

This case involves Nordic collaboration on sensitive (patient) data for personalised medicine. While the personalised medicine approach (combined with Artificial Intelligence / Machine learning analytics) is very promising in clinical research, it is still very challenging due to the diversity of data involved (often text combined with genomics and pictures) and the legal challenges related to using and aggregating sensitive personal data for secondary use. The legal and policy complexity enormously increases when data comes from different custodians subjected to different legislations. the use case proposes an alternative possible solution which does not require movement of data from their legal domain but rather a mechanism to combine securely the results of analysis computed locally.

Interviewed: Søren Brunak, PI, Professor University of Copenhagen.

Stakeholders: Cardiologists, hospitals, general health care providers and funders, geneticists, private industry within health care and pharma, general public, patients and relatives.

2.2.1 Project description

PM Heart is one of seven projects co-funded under The Nordic Council of Ministers Innovation Fund along with the Innovation Funds in the participating countries.

The project is focused on ischemic heart diseases and aims to identify the patient profiles who are at a risk of developing disease, thus enabling health care professionals to initiate profylactic treatment. The project will also try to identify those patient profiles that are not at risk, with the aim of avoiding overtreatment and terminate treatment for low-risk groups. The project focus on longer term trajectories of health and sickness as well as comorbidity.

The ambition of the project is to 1) develop algorithmic prediction for patient trajectories, and 2) improve the training of the algorithms using datasets from multiple countries.

The project started in 2019 and will run for four years. The main partners are University of Copenhagen, The Hunt Research Centre of the Norwegian University for Science and Technology (NTNU) and deCODE Genetics, a private enterprise in Iceland. The project reached out to the Swedish cardiology community and to Finnish stakeholders. The requirement of the Finnish innovation fund Sitra to have 10% cofunding from private industries, could not be met within the timeframe of the application.

2.2.2 Data

The project utilizes data from 120.000 patients in the Copenhagen Hospital Biobank. Blood samples are sequenced and sent to deCODE in Iceland. The sequenced data is highly sensitive. The data is anonymized before being sent to Iceland. The Danish genotype data are strengthened using the Decode reference genomes, thus allowing phenotype traits to be identified and algorithms to be developed. On the other hand DeCode gets insight into Ischemic heart disease, comorbidity and trajectories, which can be helpful in an Icelandic context.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

The Data is owned by the Danish National Hospital. The database has been created for internal quality processes, not research. This means that patients have not a priori given permission for research use of the data. For access, the project has obtained permission from The Danish National Health Research Ethics Committee. The data is only directly used in Denmark; anonymised analysis derived from the data is shared with Decode on Iceland.

The algorithm which is developed and benchmarked using Danish data is tested for validity in Norway using Norwegian data from the Hunt Database, owned and run by the Hunt Research Centre, under NTNU. This presents several challenges. Since the Hunt database is not a clinical but a population database, data is not identical and to some extent more limited. On the other hand the Copenhagen Hospital Biobank derives its data from Rigshospitalet - the largest and most specialised hospital in Denmark, meaning that much of the data in the database is from the most severe patient cases referred from other hospitals around the country.

Both the Copenhagen Hospital Biobank and the Hunt Database contain sensitive patient data. The datasets are curated and permanently stored with the institutions owning them. The processing of the data in the project is training for machine learning algorithms. The software used is locally developed, based on well-known, open-source machine learning packages. In addition, blood sample data is genotyped by Decode in Iceland and compared against the Icelandic genome for isolation of features.

One comment from the interviewee that it is a big challenge for small countries to provide enough cases individually to a study to have statistical relevance, particularly within genomics. Better tools and algorithms could be developed if core data could be shared across the Nordics. In this case, it would be beneficial if the Danish and Norwegian data could be aggregated rather than the algorithm used independently on the two datasets for training.

2.3 Case NorMedPregCCC

As the previous use-case, this case involves research on sensitive medical records from different countries. However, the designed approach is radically different as it aims at finding the legal framework to enable the movement of data away from the custodians and possibly across geographical borders. This case seeks to provide access to data from one country to another, arguing that GDPR does not allow data controllers to discriminate between foreign scientists from domestic scientists.

Interviewed by e-mail: Gard PO. Sundby Thomassen, University of Olso (UiO).

2.3.1 Project description

The study is a population-based registry linkage study that aims to investigate the association between maternal use of antibiotics in pregnancy and the risk of childhood cancer. The ultimate goal is to determine which antibiotics are safe to use in pregnancy when it comes to future risk of cancer in the child.

Given the low amount of cases, the data available in each country are too little to provide statistically meaningful results, therefore the need to combine the data into a single dataset. The challenge is to prepare the legal and technical risk assessment to actually move data across borders. The plan is to store the data from different sources in Norway, at the UiO operated TSD - Services for sensitive Data - infrastructure which was made for such purposes. The TSD is a platform for collecting, storing, analyzing and sharing sensitive data in compliance with the norwegian privacy regulation. The TSD is primarily an IT-platform for research bus has been used for clinical and commercial research. TSD is developed and

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

operated by UiO. The project and data acquisition has been approved in Norway, but awaits approval in Sweden, Finland and Denmark. The project is to be finished by December 2023.

2.3.2 Data

The study will use data on pregnancy complications, redeemed prescriptions, and cancer diagnoses, all of which are considered sensitive data by law. Because childhood cancer is a very rare outcome, it is necessary to combine data from health registries in Denmark, Finland, Norway, and Sweden. Data comes from health registries on approximately 2 million children, their mothers and fathers.

The project aims at sharing data on a single platform, allowing the data to be analyzed in a combined manner, as opposed to other models in which data is analyzed separately in each country and combined through meta-analysis. The idea is to make data accessible through a secure and encrypted VPN in a way that enables the PI in Norway as well as collaborators in Denmark, Finland and Sweden to analyse data from all four countries as one data file. The plan is to use TSD as a Research Infrastructure platform for testing sending data from Denmark to Norway as a testbed.

2.4 Case NICEST 2

NICEST - Nordic Infrastructure Collaboration on Earth System Tools - is a community working with climate data across the Nordics. Climate data sets are demanding in terms of volume, number of files, format and require standardization in order to be archived, discovered and re-used within the community across the region. Enabling research with aggregated data from climate modelling and observation is challenging. NICEST aims at enabling tools that enable a distributed data management and facilitate the execution of complex workflow on data distributed datasets.

Interviewed: Anne Claire Fouilloux, Department of Geosciences, University of Oslo, Hamish Struthers, Linköping University, Adil Hasan, UNINETT/SIGMA2, Ari Lukkarinen, CSC Finland.

Stakeholders: Climate scientists, meteorological institutes, universities, general public.

2.4.1 Project description

NICEST-2 - the second phase of the Nordic Collaboration on e-Infrastructures for Earth System Modelling[7] focuses on strengthening the Nordic position within climate modelling by leveraging, reinforcing and complementing ongoing initiatives. It builds on previous efforts within NICEST (a 3-year NeIC project as of 2017-01) and NordicESM (3-year NordForsk funded project from 2014-12).

2.4.2 Data

NICEST2 aims to enhance the performance and optimize and homogenize workflows so climate models such as EC-EARTH and NorESM can be run in an efficient way on future computing resources, widen the usage and expertise on evaluating Earth System Models and develop new diagnostic modules for the Nordic region within the ESMValTool, as well as create a roadmap for FAIRification of Nordic climate model data.

---

[7] https://wiki.neic.no/wiki/NICEST

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Furthermore, The NICEST2 project aims to use a variety of data sources, both public community data, EC project data, and data from individual researcher experiments. In NICEST1, all data was a subset of publicly available climate data, and all highly structured. In NICEST2, the data has become more diverse, and not all of it publicly available. For the non-public data, NICEST2 is currently investigating to understand how to best share data. Often this happens on a case-by-case basis.

2.5 Case DeepDive

The Biodiversity discipline in general, and the Nordic and Baltic biodiversity community in particular, has seen in recent years an explosion of data sets. The value of establishing a data management procedure to store and re-use these data was recognised at an early stage of this explosion and a portal was built to allow data management, data discovery and data analysis on a single entry point. The potential of such a tool has been recognised by the community and beyond. It has been suggested to extend its capability to enable management of massive data entries and facilitate cross-border analytics workflows.

Interviewed: Matthias Obst, University of Gothenburg.

Stakeholders: Biodiversity, universities, general public.

2.5.1 Project description

Over the past decade some of the Nordic-Baltic countries have developed their own Biodiversity & Ecosystem e-Infrastructures. However, due to different funding regimes, the countries have reached different levels of maturity when it comes to the technical state of the different national e-Infrastructures for biodiversity research. The goal of the DeepDive collaboration is to explore synergies in e-infrastructure development among the Nordic and Baltic countries, and establish common services based on best practice and technical interoperability to support biodiversity and ecosystem research.

2.5.2 Data

Sweden and Norway have in addition established a complementary national biodiversity information centre. The Swedish Species Information Centre and the Norwegian Biodiversity Information Centre have been collaborating closely with each other and with the respective national nodes of the Global Biodiversity Information Facility (GBIF) over the last years. In 2014, Finland started developing the Finnish Biodiversity Information Centre (FinBIF) into a national research infrastructure (RI). Norway has started the development of a Norwegian Marine Data Centre (NMDC) Research Infrastructure.

The Swedish LifeWatch is one of the most advanced biodiversity and ecosystem research infrastructure in the Nordic countries, able to access and process much of the country's biodiversity data. Norway, Finland and Denmark likewise made significant progress in assembling technical knowhow and scientific stakeholders in national LifeWatch networks. In Finland, the FinBIF portal is currently developed to become a national aggregator and RI of biodiversity data with a similar mandate as the Swedish LifeWatch. In the Baltic countries Estonia has developed several online biodiversity information solutions (complementary to the national GBIF node). Massive digitization of data from biological collections has been started in Finland and Norway, and Finland has built Digitarium – a dedicated digitization infrastructure.

Biodiversity informatics is a methodological discipline that helps biodiversity research overcome issues related to the whole value chain of data from data capture to analyses and data products regarding vocabularies, ontologies, digitization of collections, data sharing, data integration, data reliability (fitness for

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

use), data quality, visualization, analysis and long-term archival[8]. As part of maintaining the species database, the collaboration allows a system of annotations where anyone with an account can annotate data. This is effectively crowdsourcing of metadata. There is at present no system of contributors granting ownerships, access rights, or redistributor rights.

3. Recognized legal issues in the use-cases

In this chapter we review legal issues that can be identified in the use-cases described in chapter 2.

3.1. Cases PM Heart and NorMedPregCCC

The main legal issues in the use-cases described in chapter 2 come from data protection legislation, especially REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016[9] on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). This legislation applies in all European countries but it also allows some flexibility for national implementation so legislation may in some parts differ in different countries.

The main issue that needs to be observed is that when personal data information is handled, it must be clarified first who is the controller, responsible for every action of handling personal data. In use-cases where data including personal information is distributed between entities, the roles of these entities regarding the personal information needs to be defined. Are the entities sole controllers or joint controllers or are some of the entities simply processing the personal information for other entities/controllers. After that agreements need to be made accordingly as required by the legislation.

Secondly, legitimate ground for processing personal data is needed. The legislation lists what such legitimate grounds are. One is informed consent. This is usually the ground used in research projects such as the use-cases. These consents define how the controller can process data and also whether it can be shared with other organisations. If consents are not gained properly and with thought from the beginning, sharing data across borders may not be possible.

In addition, secure environments for data use are required. The data controllers need to ensure that data is secured and safe even if transferred outside their own data storage.

The regulation defines additional obligations for the controller. The controller must look after these obligations and see to that their collaborators follow them.

3.2. Case NICEST 2

The conditions for data use must be considered before data from different databases or archives, whether public or private, is accessed or gathered. Such conditions are stated in a license agreement, which is a legal arrangement between the data depositor (normally the data owner) and the repository. The agreement also specifies the obligations of the user re-using the data (data receiver), i.e., how and on what conditions the data can be re-used and shared. The license defines the framework of using the data. Without the framework specified in a licence it is difficult for the user/receiver to know how the data content may be

---

[8] https://wiki.neic.no/wiki/Biodiversity

[9] https://eur-lex.europa.eu/eli/reg/2016/679/oj

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

used. Preferably there should always be a license available; however, problems emerge if the conditions are very restrictive and do not allow sharing the data further.

In climate modelling, as well as in other disciplines that involve distributed data sets, the ownership of original data without any intellectual property rights in the database may sometimes be unclear. Regarding ownership for the kind of data referred to here, some data may not actually be considered as owned by anyone in some countries, and thus it cannot pertain to property rights. As such, it may be unclear if this kind of data can be made available without restrictions.

As previously mentioned, there may be no intellectual property rights attached to data as such, but when it is collected as a dataset or in a database, the dataset as a catalogue or database may be protected by intellectual property rights according to the national copyright legislation. The main legal base for this in Europe is the Directive 96/9/EC[10] of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases implemented in different ways in different countries.

Metadata that is provided to facilitate use or discoverability of the actual data may often include personal information. Hence, relevant obligations coming from data protection legislation need to be taken into account and followed the same way as in any situation where personal information is handled.

3.3. Case DeepDive

This case has similar kinds of legal issues as NICEST 2 as both use cases aim at collecting data from different sources and making it available.

In addition, the biodiversity data may contain information about endangered species. In this case the legislation to protect endangered species must be taken into account. In this respect at least the following directives are relevant in Europe:

Birds Directive: Directive 2009/147/EC[11] of the European Parliament and of the Council of 30 November 2009 on the conservation of wild birds (codified version of Directive 79/409/EEC as amended)

Habitats Directive: Council Directive 92/43/EEC[12] of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (Also available the consolidated version[13] of 1 January 2007 with the latest updates of the annexes)

According to the regulations not all information can be openly available and shared.

3.4. Other aspects

In addition to legislation, other aspects should be taken into account when considering legal issues.

Sometimes administrative practices and rules in an organisation acting as a controller do not allow its data (any of its data as long as it can be traced back to some of the country's person) to be sent outside the country. This means this data can only be used and shared only in that specific country. In this case it is not legislation that prevents cross-border sharing but rather the practices of an organisation.

---

[10] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31996L0009

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009L0147

[12] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31992L0043

[13] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:01992L0043-20070101

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Some of the interviewees have argued that researchers in some instances are expected to understand the legal / grant / other requirements for FAIRification of data. Interviewees have described that there may be a lack of clear guidelines, or contradictory guidelines from, e.g., funders or university management. As as result, researchers, who may face an increasing number of calls for compliance with requirements for participation in collaborations and calls for sharing data, may due to a combination of "wilful ignorance", lack of time, and a lack of expertise, choose actions contradicting the guidelines or legislation in place.

Further, some of the interviewees have argued that in relation to data sharing, participation in international projects, and regarding license issues, the legal expertise available to assist researchers at, i.e., a university may not be sufficient for handling the detailed issues. As a result, based on the use cases presented in this text, researchers may face a barrier in terms of a lack of available expertise within their home institutions.

These aspects are further discussed in the next chapter.

4. Effects of the recognized issues to sharing data across borders

EOSC, however implemented, is to function as a gateway to data, IT-services, scientific tools etc. As described above, researchers and communities interviewed from the use cases presented in this text experience barriers which are not solely legal but related to legislation, and that may arise in connection with their potential participation in EOSC. This section aims to show issues encountered by the researchers and communities interviewed. Although not exhaustive, this section illustrates some barriers which might impact uptake of EOSC among research communities presented in this text.

4.1. Effects of legislation

4.1.1. Protection of personal data

In PM Heart use-case the main legal aspect is the data protection legislation. In the interviews it was brought up that the data cannot be transferred between the Nordic countries. That was also one of the challenges for NorMedPegCCC. The administrative practice with data controllers in Denmark (Statistics Denmark / Sundhedsstyrelsen) does not allow Danish data (any of their data as long as it can be traced back to some DK person) to be sent outside Denmark. As a result, all data in a cross-border collaboration must remain in Denmark.

However, data protection legislation is similar across Europe, and legislation should not prevent transferring data between European countries. The reason for not being able to transfer data is often lack of appropriate consents. It also has to be considered that the data protection legislation does not allow indefinite processing of personal data. Processing can be done only as long as the purpose for which it is collected still exists. Thus, contemporary legislation must be taken into account already during the planning of the process, enabling data sharing and storage.

A major goal for GDPR was to enable cross-border research and make it easier to collaborate inside the boundaries of GDPR regulations and EU/EEC borders. Such cross-border will potentially have an impact on health research as it may allow for greater cohorts, particularly on rare diseases and other subjects with limited data available on national levels. The Nordics and Baltics have high quality registry data. Accessing such registry data can be very useful and even crucial for research projects. Interviewees expressed a strong

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

desire to be able to pool data across multiple countries to gain a larger cohort and a do more complete analysis. This is especially the case in instances where national datasets are small.

The use-cases describe situations where silo treatment has a negative impact on scientific results - no cross-fertilization of data. There is also a severe impact on ability to learn from and eventually treat small-sample diseases when data cannot be shared across borders.

The European data protection regulation also gives countries the possibility to some extent make their own decisions in national legislation. In PM Heart this was brought up concerning the information about dead people. In Denmark it has been decided that the data protection of personal data extends ten years after death. This kind of national consideration may create barriers for cross-border collaboration on otherwise similar data.

In NorMedPegCCC it was also observed that from a GDPR legal perspective, a research infrastructure platform can have the security measures needed to ensure compliance with the conditions of the data owner (Statistics DK / Danish Health Data Authority) while meeting the requirements of the researchers. From experience, it is not as easy in practise to allow sensitive data to flow across borders, regardless of merit and security measures. Data controllers are under heavy scrutiny and accountability towards the general public and political system, in particular when it comes to health data. This creates a strong inclination towards a conservative access regime. Or put in another way, the risk for the data controllers of releasing data across borders outweighs potential benefits. There are no formal legal requirements to provide sensitive data across borders. Data controllers have full authority to make judgement calls regarding access.

4.1.2. Data and databases

In all use-cases one aspect from the legal point of view is the use of databases. If the ownership is clear, the basic need is consent from the database owner to use the database. However, in Europe the database directive gives flexibility for countries to implement the directive. One aspect of interest is the possibility to define that the ownership of a database goes to the employer, or to leave this ruling out of the national legislation. If there is no automatic transfer of rights in employment, the employer must take care that it has the right to use the database according to its needs.

However, the conditions above are not a hindrance for data transfers between countries. The implication is that lack of thorough attention regarding data ownership may cause insurmountable obstacles. In cases where national legislation allows the employees to own database rights, institutions must take care that rights are properly transferred. This challenge is recognised for example at Finnish universities because of a national ruling[14] of the database directive concerning universities.

An issue encountered in DeepDive was also the ownership of data. This may differ and be unclear, e.g., if a Swedish researcher as a part of his job transfers data in a portal the university owns the data. If a citizen-scientist inserts data, it may be unclear who owns the data. Also, with many contributors, a database may have many owners, potentially complicating sharing.

4.1.3. Licenses

Licenses are used for both databases and software to define conditions, terms of use, etc. If licenses are not available, the conditions for use of database content and the software are not known. A database always

---

[14] https://www.finlex.fi/fi/laki/ajantasa/1961/19610404

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

has an owner. If protected by legislation, the owner has the right to determine how the database is used. Therefore, licenses, statements on conditions for availability of the database, are useful for ensuring that it is understood by the user if conditions allow the data to be shared.

If licenses are used, the challenge may be that datasets may be available under different conditions. As a result, the users are required to carefully check the conditions, especially to ascertain whether all the collected data can be shared, and if so, if it can be shared under the same conditions.

Software may be protected or made available as open source. If made available as open source, the challenges are similar to those of licenses for databases. It must be checked that the licenses used are compatible. Protected software can rarely be shared, the idea of protection being that the owner wants to decide how the software is used and make profit.

4.1.4. Case DeepDive

In DeepDive the biodiversity data faces the same challenges as previously described. Directives may allow flexibility for national legislation which then needs to be checked. As the purpose is to preserve and protect endangered species, this needs to be kept in mind when sharing data, much the same way as with personal data.

The interviews brought up some additional specific issues. The community pointed out that the legislation for red listed[15] species differs throughout the region. This result in different regulations regarding access. Furthermore, there may be different national legislation regarding obligations to report information on red listed species to national portals (eg. Artdatabanken[16] in Sweden).

For some species, such as fish, there are also commercial values that may warrant stringent demands for access. How this can be implemented may differ throughout the region.

4.2. Other aspects

As previously mentioned, legislation may not be the sole hindrance for sharing data across borders. Other aspects of importance are described and discussed below.

EOSC will be associated with terms and conditions, such as the EOSC Rules of Participation (RoP), and terms stipulated within the forthcoming EOSC statutes. This may include, regarding the RoP, terms and conditions related to data sharing, services, rules for operators, and general terms and conditions.

The future EOSC statutes, which will be the coordinating body for provisioning of services, includes, in a broad manner, formulation of agendas and setting requirements related to providing services to enable broad and secure access to data resources and data, processing services, technical development, and managing communications with researchers, communities, and service operators.

Participating in EOSC projects and in the future EOSC, researchers and communities may be expected to navigate and understand legal and other related requirements for sharing of data and scientific tools, as well as issues related to making scientific data, tools etc. available to a European, or even global, user base. Participating and navigating within this environment may place new demands on individual researchers, organisations, and communities alike.

---

[15] Red listed species are species that are under threat of extinction, see https://www.iucnredlist.org/

[16] https://www.artdatabanken.se/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

4.2.1. The researcher and community in relation to EOSC

Through the interviews conducted with researchers representing the EOSC-Nordic use-case communities, a recurring theme being raised is the barrier to participate in EOSC - sharing data, tools, et.al. across borders - due to limited capacity, time, interest, and expertise. Several of the interviewees have voiced concern that individual researchers and communities may lack the capacity and expertise to tackle issues that may arise in connection to researchers' and communities' potential participation in EOSC. As an example, the interviewees in one use case argued that correctly preparing and publishing existing research data may have a low priority for researchers when compared to generating new research results, publishing research findings, and preparing grant proposals. As a result, the motivation to spend time and effort on curating existing data may be limited.

The interviewees have in several cases argued that individual researchers' and communities' expertise may be limited to, e.g., data collection and sharing, scientific software and other tools, and development of data standards. The focus of researchers and communities within the frame of collaborations may be limited to the subject matter expertise of the individuals. Several interviewees have argued that there is a need to tackle issues outside the frame of this expertise. Examples of this may include a lack of expertise, on an individual, organisation, or community level, regarding, but not limited to, IP-rights, licence, as well as terms and conditions related to scientific software, and scientific portals and API's.

If researchers and research communities lack interest and expertise as well as time to handle practical issues related to the data sharing across the borders, there may be barriers to participate in EOSC. These are not directly legal barriers, but rather barriers related to how legal issues are handled and related the availability if support within research institutions.

4.2.2. The organisations in relation to EOSC

As described in the previous section, individual researchers and communities may be in need of institutional support at their organisations to understand and tackle requirements linked to participation in EOSC. Expectations placed on researchers and communities may in turn place increased demands on different levels of management and support functions within research organisations. Just as for researchers and communities, institutional capacity and expertise, required to investigate and tackle issues such as IP-rights, licence agreements, and legal/non-legal issues related to understanding terms and conditions for scientific software, API's, and portals, data, and resource sharing et.al. may be outside of the scope of the regular work at the organisation's management and legal departments. Lack of such capability in an organisation may act as a barrier to participation in EOSC.

Administrative practices as such may be a hindrance as well. Even though the legislation like GDPR may be flexible and allows processes to transfer data, the administrative rules of an organisation may not allow cross-border transfer. The reasons may be political or simply lack of resources to tackle legal requirements.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

5. Conclusions

Even though legislation is important and the basis for all actions, the case studies suggest that legislation is not always the most significant obstacle for cross-border data sharing.

It is clear in the use-cases that legislation create strong restrictions when sensitive personal data is involved. In these use-cases, it may in many instances not be possible to transfer databases from one country to another or doing so may be prohibitive in terms of the effort required to secure the necessary permissions.

When it comes to non-personal sensitive data, the legislation is less restrictive, allowing the sharing across borders provided the legal frameworks for the sharing is clearly formulated, for example through a license agreement. The lack of a clear legal framework for the sharing of data (ex. lack of license, undefined data owners) often observed in some use-cases is at least partly due to a poor data sharing culture, lack or pre-defined policies on data ownership and rights, as well as protective politics of the data custodians (institutions and organizations) with regard to the intellectual properties. In addition, the required data curation may be seen by researchers as a significant workload with little immediate benefit for the individual researchers or team, thus reducing the motivation to put effort into data management, and making data re-use difficult not only across borders but also across communities even within a single country.

Finally, as can be seen in some use-cases, organisations may implement local policy that discourage data sharing even in cases when sharing would not be in violation of regulation, because the perceived risk of legal, political or public exposure in case a mistake is made outweighs the perceived benefit of sharing data. This situation can be exacerbated by a lack of resources or expertise in legal matters at the organisational level and may in particular be the base in situations where the organisation providing the data is not the organisation benefiting from the use-case. In a similar manner, researchers may be discouraged from sharing data by lack of perceived benefit or by lack of available expertise in legal constraints.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Appendix A: Interview guide

Data Challenges for <TEMPLATE>

<time>

<who did interview>

<who was interviewed>

<other general notes>

Project Description <projectname>

Should include:

- partners, stakeholders
- duration - temporary or permanent
- organisational placement
- funding

The Data

Should include

- what are data source - how is the data produced or entered
    - Who owns data
    - who can access the data
    - how is access granted
    - who is the end user(s)
- what type of data is stored
- I is data sensitive?
    - What rights are there for data (licensing etc)
    - is data permanent
    - is data curated and aggregated
- how is data stored
    - where is data stored
    - when and how is data moved across border
    - Purpose of cross-border capability

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

- what kind of processing takes place
    - Which services/tools do you use to process data

Legal Issues

Should include

- existing legal framework, agreements, etc - if any
- known legal issues - resolved
- known legal issues - unresolved
- “workarounds” for known, but unresolved, issues

EOSC-Nordic project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857652