EOSC-Life

EOSC-Life: Building a digital space for the life sciences

D4.2 – Report on requirements for regulatory compliance of sensitive health data and biological and medical research data management

WP4 – Policies, specifications and tools for the management of data for biological and medical research
Lead Beneficiary: ECRIN-ERIC and BBMRI-ERIC
WP leader: Jacques Demotes and Michaela T. Mayrhofer
Contributing partner(s): Lygature

Authors of this deliverable: **Serena Battaglia, Michaela T. Mayrhofer, Mihaela Matei**
Contributors: Jan-Willem Boiten, Christian Ohmann, Irene Schlünder, Ayodeji Adeniran

Contractual delivery date: **30 April 2020**
Actual delivery date: **11 May 2020**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Table of Contents

Executive Summary

This deliverable reports on the preliminary results of the mapping of national legal and ethical requirements to understand the implications for BMS RIs and their services, especially with regard to data protection, appropriate safeguards (e.g. de-identification techniques), risk-based approaches, data ownership, and conditions for sharing and re-use of health data.

A first survey was performed to inventory the national activities and initiatives promoting the reuse of health data for research purposes, with the ultimate goal to explore whether those activities can be useful for multinational projects and data transfer between countries, in light with the implementation of the EOSC, and ultimately define principles and recommendations for the future reuse of health data. Information was collected from 15 countries in Europe and contacts were established in additional countries that will be included in the mapping.

The second step of the landscape analysis will consist of semi-structured interviews with identified representatives of the national initiatives, to further understand the legal basis and specific requirements for deposit, access and reuse of data for research purposes. Moreover, with the help of the interviews, additional initiatives may be identified.

Project Objectives

With this deliverable, the project has established the necessary foundation for a successful landscape analysis of national legal and ethical requirements to understand the implications for BMS RIs and their services, especially with regard to data protection, appropriate safeguards (e.g. de-identification techniques), risk-based approaches, data ownership, and conditions for sharing and re-use of health data. This contributes to the following objectives:

a. Promote and improve pan-European policies, standards and risk analyses to ensure full regulatory compliance of data within the EOSC, especially sensitive or classified data.
b. Propose and improve standardized solutions for working with multi-source (partitioned) environments, where datasets are not directly available for integration and in many cases not even metadata is disclosed in detail.
c. Propose solutions enabling secure federated storage, access and sharing of sensitive research or health data.
d. Promote the integration and knowledge of privacy-enhancing technologies.

Detailed Report on the Deliverable

Description of Work

Research data, digital services and facilities from the BMS RIs will be findable, “immediately” accessible, interoperable, reusable and ultimately sustainable for researchers across scientific disciplines and national boundaries. This means that European scientists will have access to advanced technology platforms, samples and support services throughout the European research area and that the resulting data are available for use and reuse through the EOSC.

EOSC needs a framework for secure archiving, discovery, dissemination and analysis of human data (genomics, translational and clinical). Services for access-controlled human data will need to cater for the sharing needed in transnational collaborative research projects as well as long-term sharing. This reuse of data will increase the rate of scientific discovery and validation leading to greater impact in the scientific and medical/healthcare fields. A secure authentication and authorisation process is essential to enable the scientific community to use human data without compromising privacy and informed consent. Additionally, support is needed in the form of guidelines and compliance processes.

WP4 in particular addresses policies and specifications for the storage, processing, access, sharing and reuse of biological and medical data for research purposes, with a special focus on sensitive data. This is an essential step to enable the European Health Research and Innovation Cloud to handle health data and research data in a transnational environment.

To that end WP4 aims to:

- Promote and improve pan-European policies, standards and risk analyses to ensure full regulatory compliance of data within the EOSC, especially sensitive or classified data.
- Propose and improve standardised solutions for working with multi-source (partitioned) environments, where datasets are not directly available for integration and metadata may not even be disclosed in detail.
- Propose solutions enabling secure federated storage, access and sharing of sensitive research or health data.
- Promote the integration and knowledge of privacy-enhancing technologies.

The first objective is therefore to inventory the national activities and initiatives promoting the reuse of health data for research purposes (selected regional activities that can have potential for a national dimension could also be considered), with the ultimate goal to explore whether those activities can be useful for multinational projects and data transfer between countries, in light with the implementation of the EOSC, and ultimately define principles and recommendations for the future reuse of health data. The legal basis of those initiatives and specific access requirements must be explored in detail. Indeed, the biomedical research

infrastructures and the services in their respective fields will be affected by the Regulation (EU) 2016/679 (General Data Protection - GDPR). Moreover, the GPDR left room for national adaptations. That is why making an inventory of national requirements for data sharing and reuse will be critical and rely on guidance documents (e.g., Code of Conduct for Health Research) and implementation of appropriate safeguards (e.g. data anonymisation or pseudonymisation) and risk assessment.

During the period covered by the present deliverable, a short questionnaire was circulated amongst the national partners of BBMRI and ECRIN, as well as other partners in EOSC-Life, EOSC-Hub and other projects:

1. **Do you know any national/regional initiatives, plans or infrastructures/organisations aiming to store and reuse human personal data (from both research and healthcare) for health research purposes?**

2. **If yes to Q.1, could you provide related links/material and names of relevant involved persons that could answer specific questions on**
    a. **The legal basis of those national registries/infrastructures**
    b. **The technical as well as procedural requirements for the data sharing (format of data, type of access, sharing across borders…)**

Information was collected from 15 countries in Europe and contacts were established in additional countries that will be included in the mapping. The list of the identified initiatives (without personal contacts) is presented in Appendix 1.

In several countries a report on national readiness for EOSC and combined strategy is under preparation and the information will be integrated afterward. For some countries, specific initiatives disease-specific (e.g. cancer registries) also exist and were added to the mapping, as an example. Moreover, several countries are implementing a national strategy to share electronic health records (not listed in the appendix) as well as building national cohorts.

After a preliminary analysis, overall in Nordic countries health data is largely accessible, with recently updated rules for secondary use of data or samples, building upon a centralised system: for instance Sweden has established a centralised agency for research permission[1]).

A similar model was implemented in UK (Health Data Research UK: an NHS centralised system to collect health data and make it available to researchers [2] and Portugal is currently building a similar strategy ("Health Data Space")[3].

[1] https://thl.fi/en/web/thlfi-en/statistics/data-and-services/data-permit-authority-findata
[2] https://www.hdruk.ac.uk/
[3] https://www.spms.min-saude.pt/2019/08/public-consultation-on-the-health-data-usage-strategy/

Other countries are implementing a different strategy, to federate existing heath and research datasets, instead of building a central place: the most relevant and advanced example is the Personal Health Train system in The Netherlands (PHT)[4]: likewise the German Medical Informatics Initiative (MII)[5] was created to close the gap between research and healthcare, bringing together university hospitals, research institutions, businesses, health insurers, and patient advocacy groups to create a collaborative research framework. France is building a similar platform (Health Data Hub)[6] and is considering an accreditation system for data hubs handling sensitive data.

The landscape as described in the present deliverable is not intended to be exhaustive and has clear limitations: a large-scale survey targeting national health/research ministries, public health institutes and other relevant bodies would probably have allowed to collect more detailed information, even at regional level; however it would have required more time and resources, for the preparation of the survey, then the collection and validation of the results. Its value would also have been limited, since such a comprehensive overview would become rapidly outdated when not extensively maintained which would require a prohibitively large workload. Therefore, the group decided to privilege direct contacts, to have a first, though as comprehensive as possible, overview of the ongoing national initiatives.

Next Steps

The second step of the landscape analysis will consist of **semi-structured interviews following a pre-determined interview guide** with identified representatives of the national initiatives to further understand the legal basis and specific requirements for deposit, access and reuse of data for research purposes. Moreover, with the help of the interviews, additional initiatives may be identified.

The semi-structured interview is a qualitative data collection strategy, in which the interviewer (researcher) asks the interviewee (informant) a series of pre-determined, but open-ended questions outlined in an interview guide (see Appendix 2). This strategy also allows for scripted as well as un-scripted follow up questions. Moreover, the interview guide provides the interviewer with a set of questions that may be rephrased in the context of the interview for the purpose of making the (rationale of the) question better understandable; therewith we aim to ensure to take national specificities into account. Finally, it will be open to the researcher if the key questions are shared with the informant prior (e.g. via email) or during the interview. Interviews may be recorded, and the written-up answers will be shared with the informants prior to analysis. This will also allow the researcher to clarify certain aspects post-interview.

[4] https://www.dtls.nl/fair-data/personal-health-train/;
[5] https://www.medizininformatik-initiative.de/en/about-initiative)
[6] https://www.health-data-hub.fr/?lang=en

The information collected will also be compared with the pan-European mapping performed by the INFRAEOSC-5b projects (EOSC-Nordic, EOSC-Pillar, EOSC-Synergy, NI4OS), with whom close contacts were already established.

An in-depth analysis including further details on the methodology will be presented in the final deliverable D4.5 “Public database inventorying the national health databases and registries and describing their access procedures for reuse for research purposes” (M30).

Delivery and Schedule

The delivery is delayed:

No

Adjustments

Adjustments made:

None

Appendices

Appendix 1 – List of national initiatives

Country	Name of initiative	website
CZECH Rep	Institute of Health Information and Statistics	http://www.uzis.cz/en
FINLAND	1. Finnish Biobanks 2. Act on Secondary Use of Social and Health Care Data 3. National Genome Centre	1. https://www.biopankki.fi/en/ 2. https://stm.fi/en/secondary-use-of-health-and-social-data 3. https://stm.fi/en/genome-center
FRANCE	1. Health Data Hub 2. Genomic 2025	1. https://www.health-data-hub.fr/?lang=en 2. https://www.gouvernement.fr/sites/default/files/document/document/2016/06/22.06.2016_remise_du_rapport_dyves_levy_-_france_medecine_genomique_2025.pdf
GERMANY	1. The Information System of the Federal Health Monitoring 2. German Institute of Medical Information and Documentation (DIMDI) 3. German Centres for Health Research 4. Medical Informatics Initiative (MII) 5. National Research Data Infrastructure (NFDI)	1. http://www.gbe-bund.de 2. https://www.dimdi.de/dynamic/en/further-services/health-care-data/index.html 3. https://dzhk.de/en/resources/feasibility-explorer/ 4. https://www.medizininformatik-initiative.de/en/about-initiative 5. https://www.dfg.de/en/research_funding/programmes/nfdi/index.html
IRELAND	1. Irish Social Science Data Archive (ISSDA) 2. HIQA Catalogue of national health and social care data collections 3. Critical Path Initiative (C-Path) EU 4. National Cancer Registry of Ireland	1. https://www.ucd.ie/issda/data/ 2. https://www.hiqa.ie/areas-we-work/health-information/data-collections 3. https://c-path.eu/about/the-future-c-paths-new-european-entity/ 4. www.ncri.ie
ITALY	National policy under preparation by the Ministry of	

	Health	
NETHERLAND	1. Health-RI 2. Personal Health Train	1. https://www.health-ri.nl/ 2. https://pht.health-ri.nl/
NORWAY	1. HELSEDATA 2. The Norwegian cancer registry 3. Norwegian Service Institution for Medical Quality Registries (SKDE)	1. https://helsedata.no/ 2. https://www.kreftregisteret.no/en/General/About-the-Cancer-Registry/ 3. https://www.kvalitetsregistre.no/registeroversikt
PORTUGAL	1. Portuguese Data Strategy for Next Generation National Health Service 2. SIMSNS system of information and monitoring of NHS data (SNS data) 3. NETDIAMOND 4. National Center for Cardiac Data Collection (CNCDC) 5. National Registry of Rheumatic Diseases	1. https://www.spms.min-saude.pt/2019/08/public-consultation-on-the-health-data-usage-strategy/ 2. https://www.sns.gov.pt/monitorizacao-do-sns/ 3. https://bioinformatics.ua.pt/netdiamond 4. https://spc.pt/cncdc/#registos 5. http://reuma.pt
SCOTLAND	1. HDR-UK Scotland 2. National Safe Haven and Scottish Medical Imaging services 3. DataLoch	1. https://www.hdruk.ac.uk 2. https://www.isdscotland.org/Products-and-Services/eDRIS/ 3. https://icaird.com/
SLOVAKIA	e-Health	https://www.ezdravotnictvo.sk/en/functions
SPAIN	1. Spanish Health Informatics Society: report on Big Data strategies (not public yet) 2. Regional initiatives (e.g. Andalusian health population database)	
SWEDEN	1. register-based research 2. SNIC (Swedish National Infrastructure for Computing) 3. Genomic Medicine Sweden	1. https://www.registerforskning.se/en/# 2. https://snic.se/ 3. https://genomicmedicine.se/
SWITZERLAND	1. Swiss National Open Science Initiative	1. https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-

	2. SIB: Setting up a national infrastructure to enable biomedical data exchange 3. SBP: Swiss biobanking Platform, national coordination platform for human and non-human biobanks 4. SPHN: National initiative to promote development of personalized medicine and health in Switzerland	P/SUK_P-2/OpenScience_Strategy_v2.5_clean.pdf 2. https://www.sib.swiss/about-sib/news/10337-a-national-infrastructure-network-to-enable-secure-biomedical-data-processing-the-biomedit-project 3. https://swissbiobanking.ch/ 4. https://www.sphn.ch/en.html
UK	Health Data Research UK	https://www.hdruk.ac.uk/

Appendix 2 – Semi-structured Interview Guide

The primary goal of the key questions (bold) below is to provide a semi-structured interview guide for the interviewer. The interviewer might rephrase questions during the interview, for which the sub-questions (italic) provide further guidance. Whereas the key questions might be shared in writing to the interview partners beforehand, the sub-questions will remain solely with the interviewer.

0. Internal Working Definitions

Health research data*: all sensitive data that is (re)used for research purposes (e.g. patient, research participant data)*

Heath care data: *Data captured and processed in the context of health care (by medical staff), e.g. for prevention, treatment and diagnostics*

Research results*: results ready for use or publishing*

---

1. Specifications on Data Initiatives:

- **What is the purpose of data processing within your institution/research infrastructure?**

Own research aims and/or providing infrastructure

- **What type of data are you dealing with?**

Health care data such as claims data, imaging, genomics, EHRs; health research data such as cohorts, biobanks, clinical trials; real-world data such as through electronic devices, patient-reported outcomes, etc.

- **Where is the data originating from (source of data)?**

For example health care institutions (e.g. clinics, ambulances, etc.), health care payer, primary health research (trials, cohorts, etc.), *secondary health research, etc.*

Please provide as much details as possible.

- **In which form are the data processed?**

Anonymised, pseudonymised, fully identifiable, etc.

- **Is it currently possible or planned to link research data with health care data?**

- inside your institution

- involving TTP

- other

If yes,

- between health research data
- between health research and health care data

Please provide as much details as possible.

- **How do you inform research participants/patients about data sharing and research results?**

Please provide as much details as possible.

- **What is the (approval) process to request health data for research purposes?**

  Please provide as much details as possible.

- **Where are your data processed and stored (e.g., cloud)?**

  - on servers in the institution
  - subcontracted service providers
  - cloud

- **How is the data security assured?**

  - SOPs, quality documents
  - audits/reviews
  - certification/accreditation

- **Under what conditions and checks are these data made available?**

Please provide as much details as possible.

2. Cloud (hosting) and data security:

- **Is there a national (or regional) strategy for the creation of a national data hub or similar?**

  if yes, provide link to the document/webpage, purpose, funders, ministries/agencies involved, challenges

- **Which agencies/funders are involved? Which challenges does the initiative face?**

- **Are there important national health data initiatives in certain sub-areas of the health data spaces that are worth mentioning (e.g. genomics) extending the level of regular national registries?**

3. Legal basis:

(database)

- **Is there a special legal regime for your field for health/research data sharing (e.g. professional secrecy, biobank law)?**

- **Do you need an authorisation/accreditation or similar to be "licenced" to collect, process, store and share your data?**

- **Is there** any **authority supervising your activities and (if) what is its role?**

e.g. supervisory authority, national ethical committee (not internal, such as access committee)

- **What is the legal basis for data processing**?

e.g. consent, clinical trial regulation or other legal allowance

- **Access modalities: what are the regulatory submissions/authorisations required to reuse the data from databases for research purposes?**

Is the transnational access and transfer of your data possible? If not, why not?

Is there a legal definition of "sensitive" or "health" data in your country? (if yes, please provide definition and legal context)

- **Is there any definition of the term "health database" (or similar) in your country's legislation or any other definition?**

e.g. 2016 WMA Taipei Declaration implemented (https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/)

- **What legal and/or ethical issues do you face when making research data accessible?**

*(*relevant *only for cloud computing)*

- **Are you storing health care or health research data in the cloud?**

- **What cloud computing system do you use?**

e.g. local, regional, national, EU vs. non-EU countries, international organisation like EMBL or CERN

- **Are there any specific requirements/regulations for cloud service providers hosting health care or health research data (local, regional, national)?**

- **Are there any specific requirements/regulations for cloud service users regarding health care or health research data (local, regional, national)?**

- **Do you require that your cloud provider has a certification, accreditation, approval or similar to host your health care and/or research data?**

4. Other:

- **Do you see other reasons than the protection of privacy to consider data as "sensitive" data?**

- **Are you aware of other national or regional initiatives? Please provide name of initiative, website and contact person, if possible.**

(This should be, in all cases, the last question)

- **Is there any additional item (issue) that you would like to draw attention to?**