Federated Platform - Data Protection Concept

Version 1.7, 2022-02-09

Definitions

Federated Platform Federated Platform of BBMRI-ERIC for federated privacy-preserving availability querying for samples and donors.

Node National Nodes means an entity, not necessarily with legal capacity, designated by a Member State, that coordinates the national Biobanks and Biomolecular Resources, and links its activities with the pan- European activities of BBMRI-ERIC. Organisational Node means an entity, not necessarily with legal capacity, designated by an intergovernmental organisation that coordinates the Biobank(s) and Biomolecular Resources of the organisation, and links its activities with those of the pan-European infrastructure, BBMRI-ERIC.

data source BBMRI-ERIC Partner Biobanks participating in the **Federated Platform** as data sources available for federated querying.

1 Nature and aims

As a part of the broader ecosystem of BBMRI-ERIC tools the **Federated Platform** will support federated search for samples, querying of data sets and support for federated data analyses and data extraction/pooling for approved requests in BBMRI-ERIC Partner Biobanks (further designated as “data sources”). It has the purpose to facilitate access to quality defined data and samples from the biobanks and biomolecular resources providing sites from BBMRI-ERIC Member States. Biobanks represent extensive collections of various types of data relevant for biological, medical and health research, biological material, and expertise, building on established trust relations with the donors of the data and samples. The biobanks are responsible for ensuring quality of the biological material and data provided to the researchers. The access is provided based on the sovereign decisions of the biobanks in compliance with ethical, legal, and other relevant regulatory requirements. Due to the nature of the federated architecture, data will remain with the data sources and will only be pooled (aggregated) when needed for a specific purpose as approved by the data source (e.g., building a specific cohort, performing centralized data quality analyses, preparing the data pool for release based on an approved request). Access is based on sovereign decisions of biobanks in compliance with ethical, legal, and other relevant regulatory requirements.

This data protection concept only describes the processes for comprehensive networking between the participating biobanks and biomolecular resources providing sites within BBMRI-ERIC. Local processes already established at those sites are not affected and are not part of this data protection concept.

2 Organisational structure, cooperation partners, responsibilities

Organisational structure BBMRI-ERIC is a European research infrastructure for biobanking and biomolecular resources. It is governed by the following bodies:

- Assembly of Members
- Management Committee
- Steering Committee
- Finance Committee
- Stakeholder Forum
- Scientific and Ethical Advisory Board

BBMRI-ERICs purpose is to bring together all the main players from the biobanking field – researchers, biobankers, industry, and patients – to boost biomedical research. For this BBMRI-ERIC offers quality management services, support with ethical, legal and societal issues, and a number of online tools and software solutions. The goal of the **Federated Platform** is to speed up the data/sample discovery and access and to support ensuring quality of data being released to the requesters.

Cooperation partners The National Notes of the BBMRI-ERIC Member States with their biobanks and biomolecular resources.

Responsibilities BBMRI-ERIC has complete sovereignty on management and operations of the **Federated Platform**. The biobanks and biomolecular resources providing sites of the BBMRI-ERIC Member States are responsible for ensuring quality of the biological material and data provided to the researchers. The access is provided based on the sovereign decisions of the biobanks in compliance with ethical, legal, and other relevant regulatory requirements. The persons and institutions responsible for processing data and the data recipients within BBMRI-ERIC are based at the operators of the central components and the sites participating in BBMRI-ERIC.

Operation of the components The partner sites represented in BBMRI-ERIC are responsible for operating the local components at the data sources. They interface with the internal infrastructure of the data source using open data models and open APIs so that data can be provided by the data source through them. The central discovery components communicate with the local components at the data sources also using open APIs and data models.

Competent body BBMRI-ERIC, Neue Stiftingtalstrasse 2/B/6, 8010 Graz, Austria, is the competent body for this project.

3 IT and network architecture, data, processes and communication channels

3.1 IT and network architecture

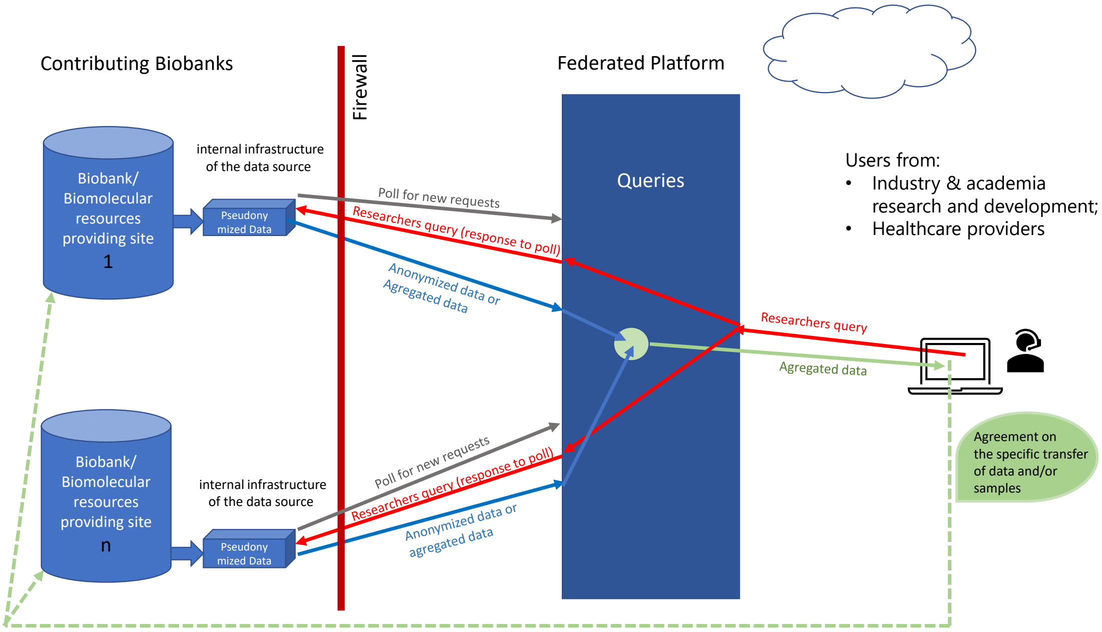
Contributing Biobanks
Firewall
Federated Platform
Queries
Users from:
• Industry & academia research and development;
• Healthcare providers
Biobank/ Biomolecular resources providing site 1
internal infrastructure of the data source
Pseudony mized Data
Poll for new requests
Researchers query (response to poll)
Anonymized data or Agregated data
Researchers query
Agregated data
Agreement on the specific transfer of data and/or samples
Biobank/ Biomolecular resources providing site n
internal infrastructure of the data source
Pseudony mized Data
Poll for new requests
Researchers query (response to poll)
Anonymized data or agregated data

Figure 1: **Federated platform** network-architecture. Note that the local components do not need any connection initiated from outside of the organization. They poll the central component hosted by BBMRI-ERIC for new requests (gray arrow) using a connection initiated from the local component; if new requests are available, they are retrieved using this connection (red arrow).

Data stay at the data sources and are technically/organizationally controlled by the data sources and data sources remain data controllers from the GDPR perspective.

Any agreement on the specific transfer of data and/or samples will be made directly between the requesting party and the biobank(s) (see Figure 1). The associated data protection aspects are to be clarified by the parties involved if necessary.

- **Federated Platform** is operated on IT infrastructure within the European data protection area. BBMRI-ERIC has full access and complete sovereignty on management and operations to this platform.
- Access control to the **Federated Platform** is fully controlled by BBMRI-ERIC and integrated with BBMRI-ERIC Authentication and Authorization Infrastructure (see chapter 3.3).
- Users accessing the Federated platform must agree to **BBMRI-ERIC Acceptable Use Policy of IT services**, except for seeing aggregate totals for the Locator where the service is available also to non-authenticated users.
- Data on queries and requests issued via the **Federated Platform** stay controlled by BBMRI-ERIC.

  BBMRI-ERIC National/Organisational Nodes (further Nodes) are allowed to monitor registration of their national data sources and monitor the queries that are being issued to them.
- Data anonymization (e.g., implemented in aggregating query results) are reviewable and periodically reviewed by BBMRI with respect to the state of the art of data analysis.
- The **Federated Platform** can and will be scanned for vulnerabilities and otherwise tested for IT security flaws by BBMRI-ERIC and its contractors.
- A provider's support plan is available, detailing what is provided in relation to technical support for the product, including an organisational chart detailing expertise of key personnel allocated to the project to include technical experience, knowledge, and capability in the mentioned area and how escalation procedures and responsibilities for support areas within the tendering organisation relate to this chart.
- The flow of information exchange between the **Federated Platform** provider and the BBMRI-ERIC is documented.

3.2 Data

Data on donors treated and cared for have been and are being collected and processed at the participating partner sites. These data are mainly compiled from existing data processing systems (e.g., clinical workplace systems, laboratory data information systems or even tumour documentation systems) to make them available in an aggregated format via the **Federated Platform** for answering research inquiries.

Data minimization always applies when providing data within the medical data (MDAT), i.e., only the data required to answer biospecimen queries is made available. Identifying data (IDAT) include demographic data that allow unique identification of donors. IDAT is neither processed nor stored by the data source internal infrastructure for providing data to the **Federated Platform**.

3.3 Services and communication channels

The **Federated Platform** is part of and integrated with the BBMRI-ERIC IT ecosystem.

Therefore, as in any other case of using BBMRI-ERIC Services dealing with personal data or anonymized data, the usage conditions as described in the Acceptable Use Policy of BBMRI-ERIC Services (see AUP in Appendix) must be accepted by any authenticated User, whereas the AUP may evolve in time to protect against additional risks. Furthermore the user agrees that any publications need to acknowledge use of BBMRI's **Federated Platform** or tools (Directory, Negotiator), and agrees with the reporting back of research results (e.g., OMICs data) to the **Federated Platform** and/or the sample/data source.

- **BBMRI-ERIC Directory**

  Data and sample discovery services.

  *The **Federated Platform** implements detailed search capabilities for more accurate search results. It is linked to the BBMRI-ERIC directory to continuously update the data for the data sources participating in the **Federated Platform**.*

- **BBMRI-ERIC Negotiator**

  Access negotiation and request tracking services.

  *The **Federated Platform** is integrated into the access negotiation pipeline.*

- **BBMRI-ERIC CRC-Cohort**

  Data harmonization, pooling, and provisioning of datasets for specific centralized data resources to simplify and expedite access for researchers.

  *The **Federated Platform** supports the process of data pooling and quality assurance.*

- **BBMRI-ERIC Authentication and Authorization Infrastructure (AAI)**

  Federated authentication and authorization infrastructure to enable trusted identities of BBMRI-ERIC research infrastructure users and control access to all services in the BBMRI-ERIC portfolio.

  *The **Federated Platform** is integrated with the BBMRI-ERIC AAI for authentication of its users and uses the AAI to provide information required for authorization decisions (see AAI in Appendix).*

- **MIABIS**

  Interoperability activities under the Community Standard MIABIS and the BBMRI-ERIC Interoperability Forum, focusing on collaborative standardization of open data models and open application programming interfaces (API) for effective coexistence of different solutions and implementations of services in the BBMRI-ERIC ecosystem.

- **Data quality assurance**

  Data quality assurance through the development and application of quality analysis tools (including AI tools for anomaly detection), such as the platform developed for the CRC cohort.

  *The **Federated Platform** internally supports data quality assurance mechanisms for data sources whose results are available to BBMRI-ERIC (aggregated or detailed, depending on the agreement between BBMRI-ERIC and the respective data source).*

- The **Federated Platform** is also expected to help donor-empowerment by implementing consent management and providing tools to support data sources in compliance with privacy regulations.

Decentralised searches

Decentralized search can be used to find suitable biospecimens and associated clinical data for research projects in aggregated form. As a search broker for the decentralized search, the **Federated Platform** provides an interface for the formulation of queries and manages them. It does not process personal data of donors. The personal data of users accessing the interface may be stored for logging purposes (see 5). The users are made aware of this.

For the decentralized search, the **Federated Platform** provides a search form that allows searching for all terms of the common model, which has been designed by consensus process in **Federated Platform** Task Force of BBMRI-ERIC. In addition, free text can be entered. Thus, in principle, there is no restriction on the data records to be queried. The query is initially stored in the **Federated Platform**. The internal infrastructure of the data source in the biobanks retrieves new queries from the **Federated Platform** at regular intervals and determine which data sets in the local data warehouse - and thus e.g., which sample materials in the biobank - match the search criteria. The query contents and the determined data sets can be viewed by an authorized person at each site.

The internal infrastructure of the data source in the biobanks and biomolecular resources providing sites (e.g., bridgeheads) only sends an aggregated number of records that match the search criteria so that the requester can get an initial overview. In order not to draw any conclusions about any individual patient, only data above a certain threshold value are made available. This is difficult regarding e.g., rare diseases, because here are only few patients and consequently few samples, which are then below the threshold value. To obtain these samples and data as well for research, it is planned that the **Federated Platform**, if getting the data even below the threshold value from the internal infrastructure of the data source in the biobanks, provides the opportunity of a further anonymization. In this case, the researcher is not informed which biobanks can provide the requested samples. The transfer of samples and data must then take place via BBMRI-ERIC.

In all other cases any agreement on the specific transfer of data and/or samples will be made directly between the requesting party and the biobank(s). The associated data protection aspects are to be clarified by the parties involved if necessary.

4 Organisational measures

Access for system administrators The data stored in the local data warehouses of the BBMRI-ERIC partner biobanks and biomolecular resources providing sites can generally be viewed by the administrators of the IT infrastructure used. The administrators may only access this data if it is necessary for them to perform their tasks. All administrators must be made aware of this and of their duty of confidentiality. As a rule, this should be the case anyway as part of their work at the responsible institution.

Authentication of users The **Federated Platform** is integrated with the BBMRI-ERIC AAI for authentication of its users and uses the AAI to provide information required for authorization decisions (see AAI in Appendix).

Authentication of components Access by one BBMRI-ERIC component to another via the internet only takes place following successful authentication, i.e., both the authorisation and the identity of the accessing components has been verified.

5 Technical measures

Security of stored data Each biobank site is responsible for the security of the stored data. This must be explained in the internal data protection concept of the respective biobank and biomolecular resources providing sites. Likewise, the ETL processes for populating the internal infrastructure of the data source from the source systems are subject to local data protection guidelines and are therefore not part of this data protection concept.

Security of communication The components of the **Federated Platform** are operated in a decentralised manner and communicate via the public internet. The confidentiality of communication is ensured by the following measures:

- Communication between the individual components always takes place via encrypted connections (HTTPS) with state-of-the-art encryption algorithms. The keys and certificates used for this must be created in such a way that they meet the currently recognized requirements (e.g., key length, algorithm, requirements on management of private keys).
- Firewalls ensure that the servers on which central components are operated can only be accessed via the protocols and ports required for communication with users or other components (usually HTTPS connections). Administrative access is restricted to the intranet of the respective operator.

- All communication processes between the internal infrastructure of the data source of the sites and the central components are initiated from the internal infrastructure of the data source. These can therefore be operated behind a firewall or a proxy server without being accessible via a public web address on the Internet.

Logging Researcher access to the components as well as access between the components is logged. Users are informed of this by the authentication service when they first access the system and are asked for their consent (see AUP in Appendix).

6 Safeguarding the rights of data subjects

Legal basis The sovereignty over the donors' data remains with the biobank site. Regarding the legal basis and internal site data processing, reference is made to the data protection concepts of the participating sites. These concepts assume that the donors have given their consent to the use of their biospecimens and data for the inter-site processes described here.

Data protection distinctions between BBMRI-ERIC and the individual sites Donor data are only collected and stored within the respective institution. There, it must first be checked whether local consent for the use and transfer of clinical and research data and/or biospecimens exists as a legal basis. If this is not the case, the respective state law regulations must be checked with the corresponding exceptions.

7 Appendix

- BBMRI-ERIC-AUP-IT-Services-1_3.pdf
- BBMRI-ERIC-AAI-Privacy-Policy.pdf

Document Log

Version 1.7, 2022-02-09 Glossary and formatting cleanup. Author: Petr Holub, Kurt Majcen

Version 1.6, 2022-01-26 Revisions based on concerns by the Nodes – schema of communication in Figure 1. Author: Petr Holub

Version 1.5, 2022-01-20 Final version as delviered by Petra Duhm-Harbeck after review by BBMRI-ERIC. Author: Petra Duhm-Harbeck

Version 1.4, 2021-12-09 Author: Petra Duhm-Harbeck

Version 1.3, 2021-12-07 Author: Petra Duhm-Harbeck

Version 1.2, 2021-11-29 Author: Petra Duhm-Harbeck

Version 1.1, 2021-11-24 Author: Petra Duhm-Harbeck

Version 1.0, 2021-11-21 Initial version. Author: Petra Duhm-Harbeck