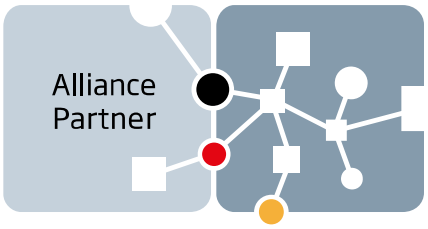
Federated Platform for German Biobank Alliance

Central Data Protection Concept

Authors	Prof. Dr. Martin Lablans, Petra Duhm-Harbeck, Pierre Delpy, Prof. Dr. Michael Hummel, Assoc. Prof. Dr. Petr Holub, Dr. Zdenka Dudová, Dr. Cäcilia Engels
Juridical person	Charité – Universitätsmedizin Berlin (German Biobank Node)

Version History

v1.2 – September 2023	Changes to address comments by PD Dr. Jörg Geiger (Univ. Würzburg) as part of his review for the TMF AG Data Protection
v1.1 – June 2023	IT infrastructure description update
	Alignment with BBMRI-ERIC DPC
	Restructuring according to TMF data protection guidelines
v1.0 – January 2018	Initial publication



German
Biobank Node
bbmri.de

Contents

1. Nature and Objectives of the Project

Given the knowledge of the complex interactions between individual genetic predispositions, lifestyle, environmental factors and molecular alterations in the development and progression of disease, biomedical research requires, in addition to the most diverse and comprehensive molecular data, data on long-term observations of responses to treatment, data on disease progression, data on individuals' lifestyle and data on environmental conditions. Biobanks play an important role in this process by systematically collecting biosamples from donors[1] and linking them to associated clinical data. Diseases resulting from individual genetic predispositions or ever smaller molecular disease subgroups can thus be adequately studied by conducting research in cross-site networks with a sufficient number of cases.

In cooperation with the German Biobank Alliance (GBA), the German Biobank Node (GBN) designed, developed and implemented modular, interoperable IT components for biobanks that enable searches across all sites – beyond the possibilities of isolated biobank management systems. A cross-location biobank network allows retrieval and compilation of biosamples and data for future (inter)national translational research projects. This network also allows German biobanks to be linked to the European Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium BBMRI-ERIC[2]. Biobanks are responsible for ensuring the quality of the biological material they make available to researchers. Access to the samples/data is usually based on the decision of the biobank use&access committee in accordance with ethical, legal, and other relevant regulatory requirements. Due to the nature of the federated architecture, data will remain with the local data sources and will only be pooled (aggregated) when agreed for a specific purpose and when approved by the data holders (e.g., to build a specific cohort, perform centralised data quality analyses, prepare the data pool for release based on an approved request) after appropriate negotiations between the data responsible person and researcher.

The GBA IT development expert site (at the DKFZ Heidelberg), together with the BBMRI-ERIC Common Service IT, has therefore developed an IT infrastructure that enables researchers to (1) find potentially available biosamples and associated clinical data for defined diseases across different sites and (2) to request these samples and data from the respective biobanks for their research. Data from the clinical documentation or collected during research projects is linked to biosamples available at the participating sites and made available for federated search queries via a local data warehouse as part of a "Bridgehead" – a local component of the system.

The local data is made available to researchers via a central search interface for feasibility studies and subsequent phases of project or collaboration preparation:

- Determine whether donor samples with the required characteristics and data are potentially available in GBN/BBMRI-ERIC-affiliated institutions in the quantity required for your research project.
- Request these medical records and biosamples for research projects using open-source tools developed by the biobanking community (BBMRI-ERIC).

Local and central components together enable findability, accessibility and interoperability of samples and associated data. The use of existing data for the Federated Platform requires appropriate patient consent. This data protection concept aims to outline how the requirements for data protection, data security, ethical and social

[1] Gender-neutral wording has been used throughout this data protection concept. "Donor" refers both to sample donors and to data donors and is used to describe persons undergoing treatment as well as test persons involved in studies.

[2] www.bbmri-eric.eu

concerns, as well as the requirements of the EU General Data Protection Regulation (GDPR) are met in terms of data flow within the Platform, while respecting the donors' personal rights.

This data protection concept applies to both the GBN/GBA and BBMRI-ERIC federated platforms for search and data analysis using (Sample) Locator. The described connections and interfaces with other IT tools run by BBMRI-ERIC (Directory, AAI, Negotiator) are necessary to cover the whole data flow from the local site to the sample delivery but is not under control of the GBN and/or DKFZ team. This workflow is based on partnership of GBN in BBMRI-ERIC as the national node (bbmri.de). The processes for the comprehensive networking of the participating biobanks within the GBN/GBA and BBMRI-ERIC are described in this document. The biobanks with the local component (Bridgehead), the central components (BBMRI-ERIC Locator and GBN Sample Locator) and the cooperative BBMRI-ERIC IT tools (Directory, Negotiator, AAI) form a federated search platform that accelerates the discovery and access to the data and samples. Local processes already established in the biobanks are not affected and are not part of this data protection concept.

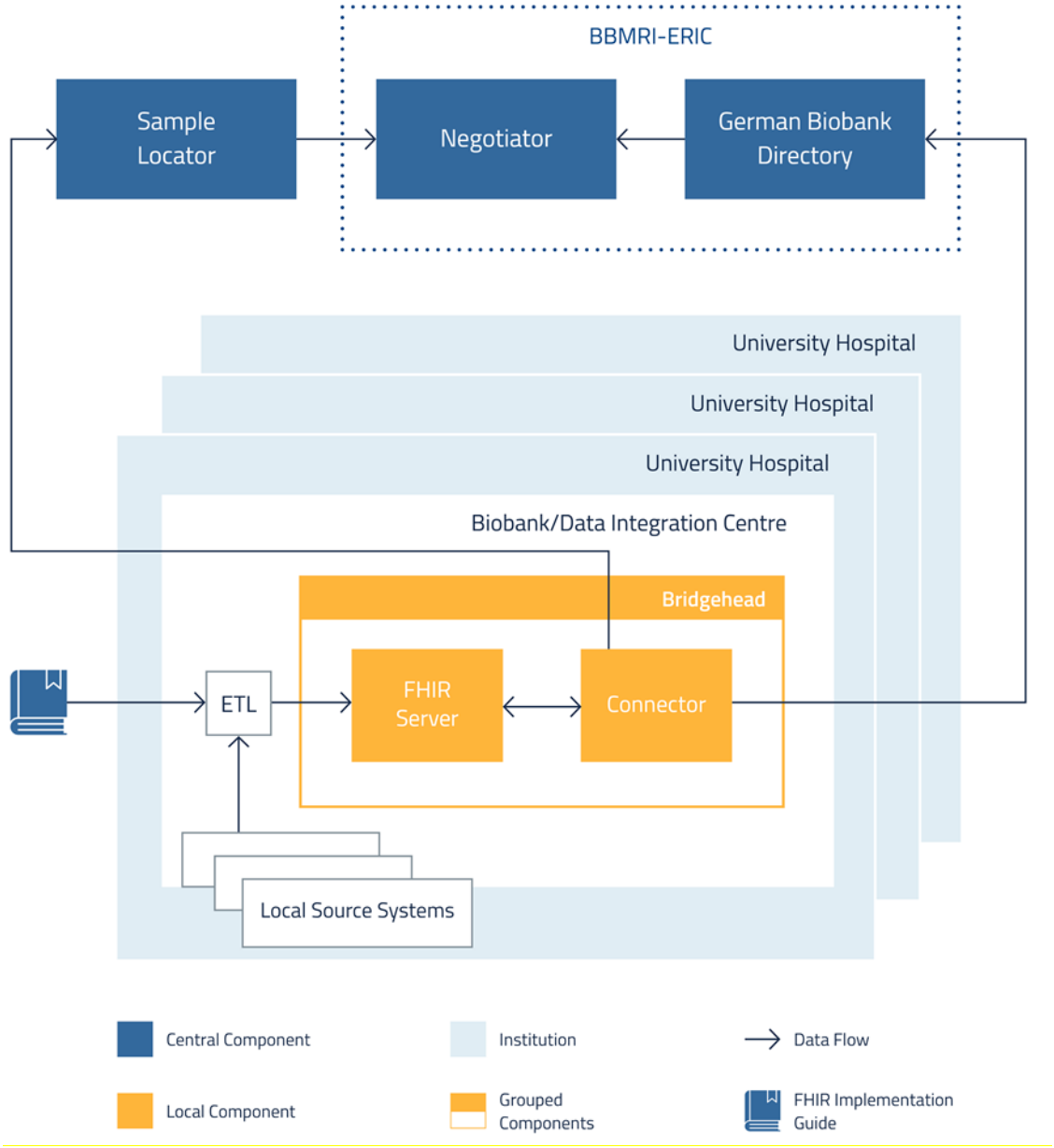


Figure 1: The underlying software IT infrastructure of the German Biobank Alliance consists of central and local components. The Extract-Transform-Load (ETL) process is based on local processes.

2. Organisational Structure

2.1 Organisational Framework

Within the institutions participating in the Federated Platform, there are local data processors responsible for the processing of personal and/or biosample data. The purpose of local data processing is, among other legal reasons, to transform the data into a harmonised and standardised form. Only highly aggregated information leaves the local components of the Federated Platform. As shown in Figure 1, the local components (distributed under a "Bridgehead" umbrella) are run within the local secure network in the university hospital and the data flow is in the direction from the local to the central component.

The administrators of the central components are responsible for further processing of the aggregates (DKFZ, see 2.3). The persons and institutions responsible for the processing of personal data or biosamples within the GBA and/or BBMRI-ERIC are located at the participating sites and at the operators of the central components.

2.2 Cooperation Partners

The sites represented in the GBA and responsible for operating the local components (called Bridgeheads), are listed at https://www.bbmri.de/ueber-gbn/german-biobank-alliance/.

2.3 Operation of the IT Infrastructure Components

Participating sites interface with the local data sources. Open data models and open APIs are used so that the data can be made available from the data source to the local component. The central components then also communicate with the local components using open APIs.

The servers for the central GBA federated search components (GBA Sample Locator for federated search and Beam operational services) are operated by the Department of Federated Information Systems at the German Cancer Research Centre (DKFZ) in Heidelberg.

The servers for the central BBMRI-ERIC services, such as the federated search components (BBMRI-ERIC Locator for federated search and Beam operational services), AAI and Negotiator, are located in the Czech Republic at CESNET[3] e-infrastructure under the maintenance of the BBMRI-ERIC Common Service IT.

2.4 Competent Body

The Charité – Universitätsmedizin Berlin has overall responsibility as the Competent Body for GBN and GBA in Germany.

The BBMRI-ERIC Headquarters in Graz has overall responsibility as the Competent Body for BBMRI-ERIC in Europe.

3. General Conditions relevant to Data Protection

3.1 Legal Basis

The use of clinical data for research purposes is based on the consent given by the respective patients.

3.2 Scope of Data Processing

Data is processed in three main scenarios: data loading, federated search and possible upload of highly aggregated data to the Directory.

[3] https://www.cesnet.cz/cesnet/?lang=en

Load Data

The local Bridgehead contains its own data store as the source of the information provided to the Federated Platform. This store is populated with the required clinical and sample data available (see the GBA Dataset in Annex) at the site. The site is responsible for setting up an extraction, transformation and loading (ETL) pipeline. The processing of patient data in this local pipeline is not part of this data protection concept.

Search

The search interface does not allow researchers to select arbitrary search criteria, but provides curated search parameters for querying. This reduces the risk of potentially malicious search queries. When a user initiates a search from the central components, each site receives a copy of the search terms. Each site performs the search locally. The search results are aggregated locally into counts of patients, counts of samples and counts of diagnoses. These counts are obfuscated[4] and then returned to the central components, where they are presented to the user. Particularly vulnerable to re-identification are results with small numbers of patients. Therefore, local components return altered patient and sample counts as follows: Zeros are reported without obfuscation; values below or equal to 10 (and not zero) are reported as 10; values above 10 are statistically perturbed and then rounded to the nearest 10.

Directory Upload

It is possible to use data uploaded to the Bridgehead to update specific highly aggregated information in the BBMRI-ERIC Directory.

3.3 Collection of Personal Data and/or Data on Biosamples

The data subject to data protection can be divided into the Medical Data (MDAT) and Identifying Data (IDAT). This nomenclature is in accordance with the TMF Guidelines on Data Protection in Medical Research Projects.[5] Identifying Data (IDAT) are demographic data that allow the unique identification of donors. IDAT is neither processed nor stored in the local component. Only MDAT is collected, including:

- biosample data, e.g., the sample type, or information on whether the sample has already been used in other projects for specific analyses;
- clinical data, e.g., coded diagnoses and tumour classifications in the case of cancer diseases, laboratory data, information on treatment and treatment outcomes, etc.;
- organisational information about the sample donors (e.g. participation in a study);
- sufficiently simplified demographic data (e.g. gender and year of birth).

The MDAT is compiled from various source systems (e.g., clinical workstation systems, laboratory data information systems, tumour documentation systems etc.) in order to make them available in an aggregated format via the GBA and/or BBMRI-ERIC central components for the purpose of responding to research queries. The principle of data economy will always be applied to the provision of data within the MDAT dataset, i.e. the loaded data will need to be stored in a different format to support querying via the central components.

[4] Obfuscating we define as employing algorithms for statistical disclosure control (SDC), to ensure that no sample donor is identifiable from the release of the central components' data. More information at https://github.com/samply/laplace

[5] "Leitfaden zum Datenschutz in medizinischen Forschungsprojekten", Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V. (www.tmf-ev.de/Publikationen/www.tmf-ev.de/datenschutz-leitfaden)

3.4 Data Integration and Storage of Personal Data

The pseudonymised data on biosamples and donors will remain stored in the local component and as such under local sovereignty as specified in the patient/donor consent. If the data can no longer be used in the intended form (e.g., if GBA and BBMRI-ERIC cease to exist), it is the responsibility of each individual site to either anonymize or delete the data, in accordance with the given patient consent used as legal basis for data storage.

4. Rights of the Data Subjects

Locally pseudonymised records stored in the Bridgehead can be queried via the Federated Platform and only aggregated counts and stratification leave the biobank site. In both cases, the response to the query complies with the state and federal data protection regulations applicable at the site. The data described is protected by comprehensible access rights. The necessary components and processes are described in chapter 6.

4.1 Data Protection Distinctions between GBA/BBMRI-ERIC and the Individual Sites

Personalized donor data is only stored within the individual institutions and the donor data remains under the sovereignty of the site. Please refer to the privacy policies of the participating sites for information on the legal basis and internal procedures for data processing. These confirm that the donors have given their consent to the use of biosamples and data collected from them.

If this is not the case (e.g. existing samples without explicit consent), a site may still decide to import data about said sample into the local component to make it findable using the Sample Locator, since only aggregated, non-identifying data leaves the institution. However, actual transfer and/or use of the samples or associated data will still require obtaining a legal basis. It is up to the site to ensure this legal basis, usually by obtaining an informed consent.

4.2 Information about Stored Data

All donors whose data is used in the technical components have the right to receive information about their stored data. Requests for information must be made in writing to the treating clinic/responsible institution.

4.3 Deletion and Right to be Forgotten

All donors whose data is stored and used in the technical components have the right to withdraw their consent to the processing of their data in the GBA/BBMRI-ERIC at any time. The data in the Bridgehead will be deleted accordingly by the administrator responsible for the data in the Bridgehead of the biobank concerned. The revocation must be sent to the location where the biosamples and associated data are stored.

The responsible operators at each site must delete the data from the Bridgehead within a reasonable period of time and notify the donor of the deletion. Deletion within data backups, which is usually not possible, is not required if the backups can only be viewed by the responsible system administrator and outdated backups are regularly deleted. In the unlikely event that data needs to be restored from a data backup, each site ensures that data from the patients who have withdrawn their consent is not restored to the Bridgehead.

5. Organisational Measures

5.1 Access to Data by System administrators

The data stored in the Bridgeheads' local data warehouses can be viewed by the Bridgehead's IT administrators. Administrators are only allowed to access this data if it is essential for the performance of their duties. 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 employment with their responsible institutions.

5.2 Access to Data by Biobank Administrators

Biobank administrators have access to information about samples and may also have access to personally identifying information, for example names or e-mail addresses of users. At some sites, depending on how data integration is implemented, biobank administrators may become privy to identifying data of sample donors as part of a record linkage process; this data is not stored in the Bridgehead. Biobank administrators are bound by the legal obligations of their institution to protect this information.

5.3 Access to Data by Users of the Search Infrastructure

Users of the central search infrastructure do not have direct access to information about samples or patients. They can only see aggregated numbers of patients and samples that match their queries. They may choose to contact the sites holding the samples and request further information. However, this process is outside the scope of this data protection concept.

5.4 Use of Personal and Biosample Data

Informed consent, in which the donors are informed of their right to information and withdrawal, is the legal basis for data processing and the use and transfer of biosamples. Due to the multitude of national and international projects depositing samples in partner biobanks, it would be impossible to enforce one uniform consent text. However, we provide a consent template (see Annexe) building on the expertise of recognized actors and organizations throughout Germany, such as the Biobanks Working Group of the Medical Ethics Committee and the Data Protection Working Group of the Technology and Methods Platform for Networked Medical Research. Sites that have implemented electronic recording of the consent status of their donors can transfer this consent status in the ETL process when filling the Bridgehead.

All necessary details regarding informed consent and the legal basis for the use of biosamples and patient data are fully on the biobank's side. The IT infrastructure provided by the GBA/BBMRI-ERIC should comply with the above legal requirements.

5.5 Tracking Activity within the Infrastructure by Logs

Access by researchers to the components, as well as access between the components, is logged. Users are informed of this fact when they go through the authentication service via the "Terms of Use"[6]. The first time they access the federated system, their consent is requested. The log includes at least the following:

- the identity of the person or component accessing it
- the date and time of access
- the content of the access (data transmitted, aggregated where appropriate) or information from which it can be reconstructed (e.g. reference to a database entry or similar).

The log is stored on the server on which the Service is running, together with the corresponding server payload. The log files are deleted after twelve months. The data collected will only be processed and viewed for the following purposes:

- technical administration (in particular troubleshooting)
- detection of possible misuse
- generation of anonymous usage statistics

[6] DKFZ-Datenschutzerklärung - https://www.dkfz.de/de/datenschutzerklaerung.html

5.6 Data Quality Assurance

Bridgehead administrators are responsible for ensuring a high level of data quality. To this end, GBN requests them upon first joining the Sample Locator and in regular intervals (1-2 years) to generate so-called technical quality reports by clicking a corresponding button on the Bridgehead's local user interface. The Bridgehead will then process all data in its internal data warehouse and generate a report containing aggregated statistics, with highlighted irregularities and deviations of the data from the given data specification. The Bridgehead administrator uses this report to correct errors in their ETL processes. Optionally, they may also share the report (completely or in parts) with GBN to request assistance.

In addition, GBN may execute automatic plausibility queries over the Sample Locator's query infrastructure. Any irregularities are then reported to the Bridgehead administrator, who can locate the offending datasets locally and perform any corrections. Plausibility queries are performed in the same way as any other query to ensure no additional security issues are introduced.

6. Technical Measures

When processing personal data, the protection objectives set out in Article 32 (1) of the GDPR, such as confidentiality, integrity and availability of the systems and services, and their resilience in relation to the nature, scope, circumstances and purpose of the processing operations, are taken into account and any existing risks (DPIA - Data Protection Impact Assessment) are mitigated by appropriate technical and organisational measures. This section describes the general infrastructure and technical measures to protect the data collected.

6.1 System Components

Local Components

The local component of the Federated Platform, the Bridgehead, must be established at participating site. It consists of hardware and software components that are installed and operated independently by each site. The security of locally stored data is therefore the responsibility of each institute/site and is described in their local documentation. Similarly, the data integration processes for delivering data from the source systems to the Bridgehead are subject to local data protection guidelines and are not part of this data protection concept (**Security of locally stored data**). In order to keep the Bridgehead updated and correctly configured, it is centrally updated and monitored (see 6.3).

In terms of functionality, the Bridgehead's data store serves as a federated data warehouse for a site's data connecting the site and their data to the central components. The features of the Bridgehead are:

- *Harmonised data storage*: Data is stored within the Bridgehead in a standardised format so that it can be understood by the central components of GBA and BBMRI-ERIC.
- *Visibility to GBA/BBMRI-ERIC*: Aggregated results are provided in response to researcher queries during federated searches through the central component.
- *Data sovereignty compliance*: The Bridgehead allows sites to participate in federated searches without having to upload donor-related data to an external entity. This supports data privacy and sovereignty.

The Bridgehead consists of several software components, which are installed locally in the biobanks[7]. These components of the Bridgehead are controlled by the respective site – sovereignty over the data stored in these components therefore remains with the institution where it was collected.

[7] https://github.com/samply/bridgehead

Federated Search

Federated searching enables the detection of appropriate biosamples and associated clinical data for research projects. The centrally operated federated search component provides a query formulation and a management interface. Since it only receives aggregated data, it does not process personal donor's information. However, the personal data of users accessing the interface may be stored for logging purposes (see 5.5). The query created by the user is first stored in the central components. The local components in the biobanks periodically retrieve new queries from the central components using the Beam infrastructure (see next paragraph). They then run the query internally against the data in the local data warehouse and determine which records – and thus also which samples in the biobank – match the search criteria. Locally, the content of the query and the records identified can be reviewed by an authorised person at each site. Aggregated information about records matching the search criteria is returned to the central component and is visible to the querying user.

Beam

Samply.Beam is a distributed task broker designed for efficient communication across stringent network environments. It provides the most commonly used communication patterns across strict network boundaries, end-to-end encryption and signatures, as well as certificate management and validation. It is used for the communication between central components and individual biobanks, and allows biobanks to register with the search infrastructure in a secure manner.

GitLab

The DKFZ operates a central GitLab instance that stores the necessary configurations for each site connected to the GBA Federated Platform.

In the case of the BBMRI-ERIC Federated Platform, the central GitLab instance is operated by BBMRI-ERIC CS IT administrators.

6.2 System Model

Data

Technically, the sites randomize the aggregate patient/sample counts using methods of "statistical disclosure control". Specifically, a random value drawn from a Laplace distribution with mean 0 and a standard deviation of 5 is added to the patient/sample count. Since this random value can be positive or negative, an upward or downward deviation from the real patient/sample count is possible. This procedure strongly resembles the concept known as "differential privacy"[8], although not all of its rigorous mathematical statements can be achieved in practice. Therefore, a further obfuscation step is performed in the form of rounding to the nearest 10th digit. As also described in the statement of the Art.29 group[9], the greatest risk of re-identification in differential privacy is posed by performing multiple queries, therefore a) randomization is saved with respect to the query result and b) user control and restriction measures, described in the following, are implemented. The data of donors who are treated and cared for at the participating partner sites is only collected and processed only there and it is not part of this data protection concept.

[8] Dwork D. Differential Privacy, ICALP 2006. doi: 10.1007/11787006_1

[9] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf, retrieved 06/17/2023, pp. 17-19

Importing Data into Bridgehead

Data from a site's various source systems is imported into the local Bridgehead during the extract, transform and load (ETL) process. The ETL process imports pseudonymised data into the local Bridgehead. The data is stored in the part of the Bridgehead that acts as the local data warehouse.

BBMRI-ERIC Negotiator

Once the user making the request trough the central components of the federated system has identified the biobank(s) with the required samples, the query can be sent to the BBMRI-ERIC Negotiator. This tool accelerates the communication between the requester and the biobank(s) and tracks the requests. All necessary information about the project for which the samples are needed can be easily exchanged between the requester and the biobank(s). All users of the Negotiator tool use the same authentication and authorisation infrastructure (AAI) that is used to log in to the central components. Any agreement on the terms and conditions for the transfer of data and/or samples of interest resulting from the communication between the requester and the biobank(s) will then be made directly between the parties involved. Data protection aspects related to this agreement will need to be clarified by the parties involved, if necessary.

As soon as a researcher requests samples using the Negotiator tool, the biobanker who receives the request from the researcher will see a button redirecting to the Bridgehead of the site. This button must be configured in the Bridgehead by the Bridgehead administrator. It can be used to visualise the identified samples from the query performed in the Locator.

BBMRI-ERIC Directory Synchronisation Tool

The BBMRI-ERIC Directory is a tool that collects and makes available information on biobanks across Europe. It provides a publicly accessible central listing of biobanks and their collections in the BBMRI-ERIC member states. The Bridgehead can be linked to the BBMRI-ERIC Directory to update the data for the data sources participating in the Federated Platform. The activation of this tool is optional. Once activated, basic information from the Bridgehead can be updated in the Directory on a regular basis.

6.3 System Operations: Software Updates & Monitoring

By default, Bridgeheads are subscribed to a centrally managed monitoring service operated by DKFZ (for GBA) and BBMRI-ERIC CS IT (for BBMRI-ERIC Federated Platform). This service will receive information on which Bridgehead components are functional and will automatically run test queries to ensure that the infrastructure is working as intended. The test queries are subject to the same restrictions and security mechanisms as any other Locator query; in particular, it is impossible to gain further insight into local data. Similar to the central federated search component, no patient data is processed by the central monitoring services. Monitoring results are visible to central service administrators. Site administrators can opt to receive emails when monitoring detects problems at their site.

Similarly, in order to quickly react to potentially outdated / misconfiguration and security issues due to outdated software, updates to the Bridgehead software and configuration changes are by default automated, in which case they are pulled from central repositories hosted at DKFZ Heidelberg or BBMRI-ERIC CS IT and automatically applied locally. Any site admin can review the respective configuration before/after as well as monitoring data at any time, and also – although not recommended – opt out of one or both mechanisms.

6.4 Authentication and Authorisation Infrastructure (AAI)

User authentication for central GBA components is performed by the "Life Science AAI" central authentication service, maintained by the operators of the Life Science RI dedicated for GBN/GBA and BBMRI-ERIC services. This

federated authentication and authorisation infrastructure enables identity verification of users of the federated infrastructure and controls access to all services in the BBMRI-ERIC portfolio. Users accessing the BBMRI-ERIC and GBA IT infrastructure must agree to the Life Science AAI Acceptable Use Policy[10] and BBMRI-ERIC Acceptable Use Policy for IT services[11]. The AAI provides a unified single sign-on, using the same credentials everywhere. Using federated authentication protocols (SAML, OpenID Connect), user authentication is delegated to the identity provider at each site, who meets the eduGAIN criteria[12]. Alternatively, Life Science RI is offered to users without institutional identity providers; this login meets the REFEDS Single Factor Authentication profile[13]. The sites themselves are responsible for authenticating users to use local components.

6.5 Network Security (IT security)

The components are federated and communicate via the public Internet. The confidentiality of the communication is ensured by the following measures (**Communication security**):

- Communication between individual components of the federated IT infrastructure is always over encrypted connections (HTTPS). Keys and certificates used for this purpose are created in such a way that they meet the currently accepted requirements (e.g. key length, algorithm), and are issued by certification authorities meeting the Mozilla Root Store Policy[14].
- Firewalls ensure that the servers running the central components are accessible only through the protocols and ports required to communicate with users or other components (generally HTTPS connections). Administrative access is restricted to the people nominated by the operator.
- All ingress communication towards the sites' Bridgeheads is mediated through the Beam component of the Bridgehead. Thanks to Beam's architecture, Bridgeheads can therefore operate behind firewalls and proxy servers without being accessible via a public web address on the Internet.
- There is no direct communication between local components in different institutions.
- The connection between the Negotiator and the local component (see 6.2) is only available when using a virtual private network (VPN) and logging into the Bridgehead with the appropriate role/rights.
- Patient and sample counts are obfuscated locally before being sent to the central infrastructure. This is done by introducing some randomness into the count and then rounding it to the nearest multiple of ten.

[10] https://lifescience-ri.eu/ls-login/ls-aai-aup.html
[11] https://web.bbmri-eric.eu/Policies/BBMRI-ERIC-AUP-IT-Services-1_3.pdf
[12] https://edugain.org/
[13] https://refeds.org/profile/sfa
[14] https://www.mozilla.org/en-US/about/governance/policies/security-group/certs/policy/

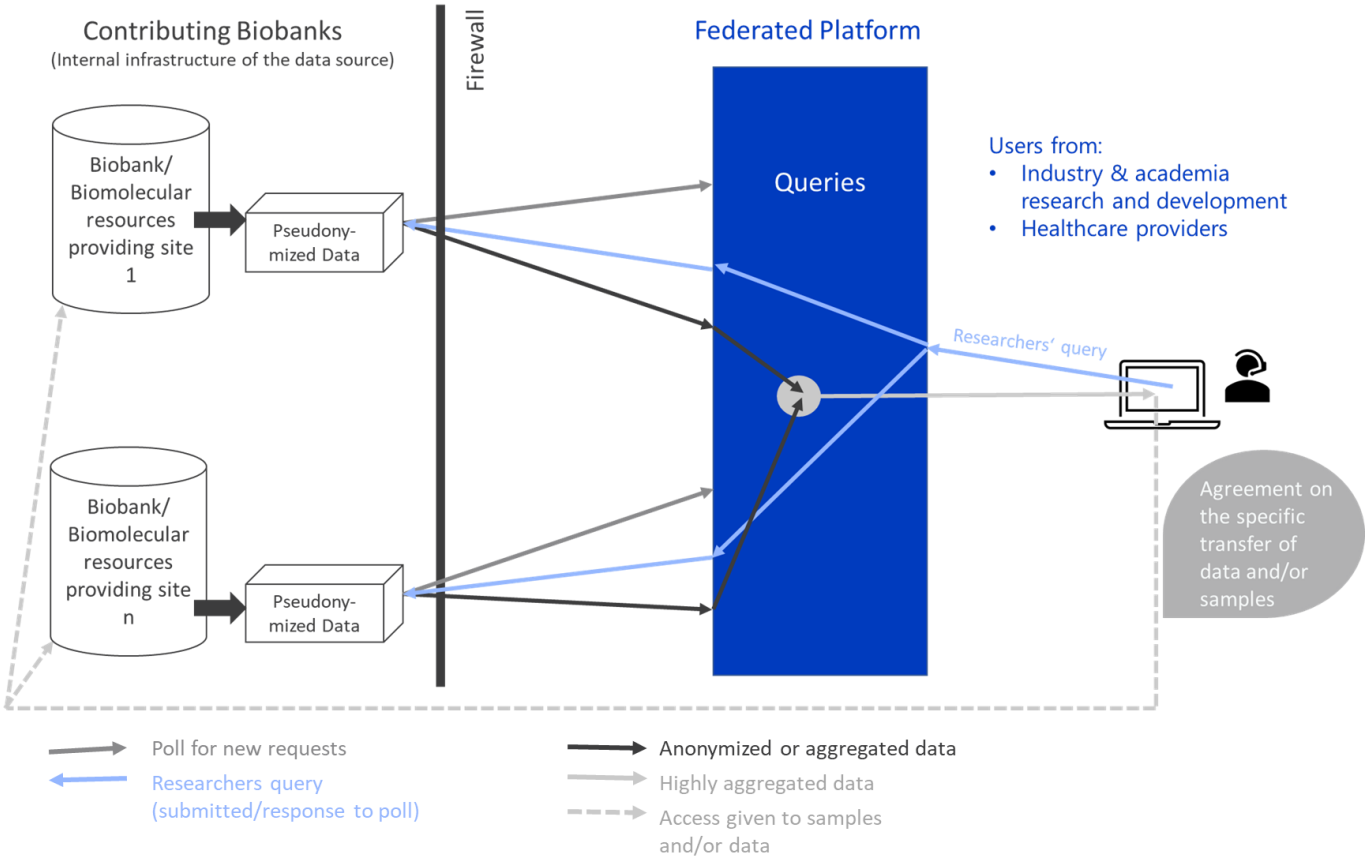


Figure 2: ***Federated Platform data flows****. The local components do not require a connection initiated from outside of the organisation. They poll the central components hosted by DKFZ or BBMRI-ERIC for new requests (**grey** arrow) using a connection initiated by the local component; if new requests are available, they are retrieved using this connection (**light blue** arrow). Anonymized and aggregated data is sent to the central component (**black** arrow) where further aggregation occurs and highly aggregated data is shown to researcher (light grey arrow). Figure adopted from BBMRI-ERIC DPC[15].*

6.6 Backup Strategy

Backups of locally stored data in the Bridgehead should be performed automatically on a regular basis. All backed up data remains onsite according to their requirements.

6.7 Used Encryption Technology

All communication with web-based components is protected by HTTPS. Ingress communication towards Bridgeheads is handled by Beam, which provides encrypted channels for all transferred data (**Transfer control**).

6.8 Pseudonymisation

Pseudonymization is a local service under the responsibility of the local institution and therefore not described here. The Bridgehead is controlled locally by each site, and a donor's identity cannot be re-identified by the biobank or IT staff.

[15] https://web.bbmri-eric.eu/Policies/FedPlatform_Data_Protection_Concept-1_7.pdf

6.9 Additional technical Measures

Logging

Access to and between the components is logged. The user is informed of this by the AAI in the form of a user agreement (AAI Privacy Policy)[16] the first time he or she accesses the system and his or her consent is obtained.

Access control - authorised

Only authorised employees of the respective operators have access to the servers. Authorisation is controlled electronically via locking systems and is automatically logged.

Access control - administrative

Administrative access to the central servers is granted by the authorised administrators of the operators of the respective central components.

Access control - Bridgehead

Access at network and application level is protected by password-based authentication mechanisms. User rights are controlled by a rights and roles concept. This ensures that each user only has access to the data they are authorised to access.

Input control

All data operations, including imports via ETL processes, modifications, and related actions, are logged in the Bridgehead. In addition, data is checked for compliance to specification upon import.

Availability control

The server rooms where the central instances are running in a data center certified with TSI.Standard 4.3 Level 1 (expanded)[17]. Server backups are performed automatically on a daily basis. Security scans are carried out regularly to protect the systems from attacks. All necessary components such as proxies and firewalls are in place and regularly updated. Only authorized personnel are involved in these processes.

7. Local environment

GBA/BBMRI-ERIC biobanks may describe specific characteristics of their site here. Where possible, reference should be made to the site's data protection concept.

[16] https://web.bbmri-eric.eu/Policies/BBMRI-ERIC-AAI-Privacy-Policy.pdf
[17] https://www.tuvit.de/fileadmin/Content/TUV_IT/zertifikate/de/66868UD.pdf

Annex

Patient consent

The model text on the donation, storage and use of biosamples and the collection, processing and use of data in biobanks which we recommend has been developed and coordinated by recognised actors and organisations in Germany, such as the Biobanks Working Group of the Medical Ethics Committee and the Data Protection Working Group of the Technology and Methods Platform for Networked Medical Research[18]. There are several language mutations of the templates and each GBA site can adapt the document to local requirements. For example version 1.7.2 includes Transfer of data and biosamples to non-EU states. The actual information about what consent form is used on the site can be obtained from the GBA site representative on demand.

GBA Dataset

The dataset used in GBA is stored on SIMPLIFIER.net.

Data Protection Impact Assessment

A data protection impact assessment (authored by Petra Duhm-Harbeck) has also been prepared for local processing, see[19].

BBMRI-ERIC Data Protection Concept

The GBA Data Protection Concept is aligned with the BBMRI-ERIC Data Protection Concept, which describes the BBMRI-ERIC Federated Platform[15]. It is not necessary to conduct a double Data Protection assessment if the biobank or institution is interested in connecting to both of the GBA and BBMRI-ERIC Federated Platforms.

Acknowledgement

This data protection concept is based on the general data protection guidelines of the TMF e.V. and the data protection concept of the GBA version 1.0 written by Martin Lablans, Dr. Esther Schmidt, Petra Duhm-Harbeck, Prof. Dr. Hans-Ulrich Prokosch, Prof. Dr. Michael Hummel. We would like to thank PD Dr. Jörg Geiger (Univ. Würzburg) for his review of this concept for the TMF-AG Data Protection, and his constructive comments that led to the improved version v1.2.

[18] https://www.medizininformatik-initiative.de/de/mustertext-zur-patienteneinwilligung

[19] https://www.bbmri.de/fileadmin/user_upload/PDFs/GBA_Datenschutz_Folgenabschaetzung_20180331.docx