

Federated Platform Call for Tender

Version 2.0

The goal of the tender is to provide BBMRI-ERIC with a **platform that 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”). This platform is further designated as a “**federated platform**”. By the nature of the federated architecture, the data should stay at the data sources by definition and only pooled (aggregated) when needed for a specific purpose after approval of the data source (e.g., building a specific cohort, doing centralized data quality analyses, preparation of data pool for release based on approved request).

The federated platform will become a part of the broader ecosystem of BBMRI-ERIC tools to facilitate access to quality defined data and samples from the biobanks from BBMRI-ERIC Member States. The biobanks are rich troves 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. *The purpose 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.*

The IT ecosystem of BBMRI-ERIC provides a portfolio of services:

- Data and sample discovery services, such as BBMRI-ERIC Directory.
The federated platform shall implement in-depth search functionality to more accurate discovery results.
- Access negotiation and request tracking services, namely BBMRI-ERIC Negotiator.
The federated platform shall integrate with the access negotiation pipeline.
- Data harmonization, pooling, and providing data sets for specific centralized data resources in order to simplify and expedite access to the researchers, such as BBMRI-ERIC CRC-Cohort.
The federated platform should support the process of data pooling and quality assurance.
- Federated authentication and authorization infrastructure to enable trusted identities of users of BBMRI-ERIC Research Infrastructure and to control access to all the services in BBMRI-ERIC portfolio, so called BBMRI-ERIC Authentication and Authorization Infrastructure (AAI).
Federated platform shall be integrated with the BBMRI-ERIC AAI for authentication of its users and use AAI to provide information necessary to make authorization decisions.

- Interoperability activities organized around MIABIS community standard and BBMRI-ERIC Interoperability Forum, focusing on community standardization of open data models and open application programming interfaces (API) for effective co-existence of different solutions and implementations of services in the BBMRI-ERIC ecosystem.
Federated platform is expected to use open APIs and data models at least in the following two components: (1) local components at data sources shall be interfacing the data source internal infrastructure using open data models and open APIs, so that the data can be provided by the data source using these, (2) the central discovery components shall communicate to the local components at the data sources using open APIs and data models, so that the local components at the data sources can talk to different central components.
- Data quality assurance by developing and applying quality analysis tools (incl. AI tools for anomaly detections), such as the platform developed for the CRC-Cohort.
The federated platform shall internally support mechanisms for data quality assurance at data sources, results of which would be available to BBMRI-ERIC (aggregated or detailed, depending on agreement between BBMRI-ERIC and the particular data source).

The platform is also expected to contribute to donor empowerment: implementing consent management and providing tools to support data sources in compliance with the data protection regulations.

Specification of the Federated Platform

The specification is provided as follows:

- Criteria are split into groups.
- For each group, there are hard requirements provided and optional/bonus features.
- Within each group, the hard requirements are sorted in on particular order.
- Optional/bonus features are provided without particular weights at the moment (weights can be assigned in the later stage of preparing the call for tender).

Qualifications of the platform

Requirements.

- The platform shall have a proven record in supporting research on an international or comprehensive national scale (e.g., international projects).

Optional/bonus features.

- Demonstration of an already running system operating on a regular basis.

Governance, data protection and business model

Requirements.

- Data shall stay at the data sources and shall be technically/organizationally controlled by the data sources and data sources remain data controllers from the GDPR perspective.
- Data on queries and requests issued via the federated platform shall stay controlled by BBMRI-ERIC.
- BBMRI-ERIC shall have complete sovereignty on management and operations of the platform.
- National nodes of the BBMRI-ERIC shall be allowed to monitor registration of their national data sources and monitor the queries that are being issued to them.
- The platform shall be operated on white label basis, i.e., the design of the user interface must be fully customizable by BBMRI-ERIC in terms of branding.
- Data anonymization (e.g., implemented in aggregating query results) shall be reviewable and has to be periodically reviewed with respect to the state of the art of data analysis.
- The platform shall allow BBMRI-ERIC to operate it in GDPR compliant manner.
- Platform provider shall support BBMRI-ERIC in undertaking an DPIA (needed at least because of the federated analysis and data pooling mechanisms requested below).
- If the federated platform is operated as a service, data loss protection shall be ensured and the data shall be stored on encrypted data storage.
- If the federated platform is operated as a service, flow of information exchange between the federated platform provider and the BBMRI-ERIC shall be documented.
- The platform shall comply with GDPR the requirement on transparency, i.e., provide extensive information to data subjects on processing of their personal data when requested.
- The federated platform shall support handling special categories of data under GDPR Art. 9 (e.g., genetic and biometric data).
- Access control to the platform has to be fully controlled by BBMRI-ERIC and integrated with BBMRI-ERIC Authentication and Authorization Infrastructure.
- The federated platform provider shall allow the system to be scanned for vulnerabilities and otherwise tested for IT security flaws.
- Details of any independent audit or certified testing of the software to industry standards, e.g. ISO 27001 or equivalent
- Clear and concise English-language documentation including technical specifications for all infrastructure requirements, details of tools and methodology used in the transmission and storage of data and how the system is secured against attackers must be provided.
- A support plan shall be produced, 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.
- Training on the platform shall be provided to BBMRI-ERIC and Common Service IT staff.

Optional/bonus features.

- Implementation of custom data anonymization strategies will be awarded a bonus.
- Traceability of all actions on the platform will be awarded a bonus.
- Details of quality management, software development and project management methodologies and any certifications held related to those areas will be awarded a bonus.
- Details of all packages installed with the system with an explanation of their function in the system should be provided.
- Software available as open source will be awarded a bonus.
- The ability and expertise in environments similar to those described in this invitation to tender in terms technologies for both custom developed systems and commercial off the shelf software will be awarded a bonus.

Data extent and quality

Requirements.

- The following data types shall be supported by the federated platform for querying: biosample related data, phenotypic data, clinical data, genomic mutation data, proteomic, transcriptomic, metabolomic, metadata on imaging.
- The data models shall be extensible by BBMRI-ERIC.
- The following matrix shall be supported for common complex diseases with at least the following elements:
 - disease classification (ICD-10 code mandatory, other ICD versions, SNOMED CT and Orpha codes optional) for primary diagnosis;
 - personal data (date of birth internally at data source for different calculations of clinically relevant events, sex, ethnicity, age at diagnosis, familial incidence of disease under study)
 - results of physical examination (blood pressure, weight, body mass index, ...)
 - lifestyle data/risk factors (smoking, alcohol, fat intake... / direct causality : any polyps, hypertension, obesity...)
 - familial history - comorbidity: Charlson index
 - performance scale: Karnofsky scale, ECOG
 - laboratory data
 - molecular testing / analysis
 - histopathology
 - metadata on imaging/instrumental data in DICOM format (X-ray, PET, Scan, MRI, colonoscopy, endoscopy...)
 - treatment/ treatment response / other treatments
 - surgery
 - follow up data/recurrence
 - survival
 - TNM staging classification and/or grading
- It shall be possible to specify specific data model for each of the diseases to be included in the database. It constitutes a minimum data set that will enable to collect samples

from patients with similar inclusion criteria that will be used for genomic and other ‘-omics’ studies.

- All the properties that are necessary for unambiguous interpretation of the results have been defined.¹ This includes the following properties for each collected variable (attribute):
 - unique identifier of the variable
 - short description (label) of the variable - to be used in forms
 - semantics - i.e., definition of meaning, including references to existing clearly defined official standards or community “standards”, including existing ontologies.²
 - syntax
 - including data type (elementary types such as boolean, float, integer, free text, specifically structured text, etc., array or lists of elementary types),
 - including coding (e.g., IEEE 754 for floats, regular expressions for structured text).
 - list of allowed units including their conversion algorithms (with “non-existent” and “unknown” interim options)
 - level: REQUIRED, OPTIONAL, RECOMMENDED³
 - relation to entities (patient, examination, etc.) - will be used for developing the formal model.
- The platform shall provide mechanisms for data quality control at the data source, including running custom data quality checks designed by BBMRI-ERIC based on the common data models provided by the federated platform.
- The platform shall support monitoring of aggregate results of the data quality control by BBMRI-ERIC.

Optional/bonus features.

- Ability to support proteomic, transcriptomic, metabolomic data will be awarded a bonus.
- The ability of the platform to provide interfaces to the data lakes (i.e., storage with data in their original formats/structures) will be awarded a bonus.
- Implementation of existing data quality checks will be awarded a bonus based on extent of existing available data quality checks. The following data checks will be considered:

¹ Each single item of the dataset is identified based upon the criteria defined. The goal is to provide a detailed description of the data structure in order to (a) create useful datasets for each disease, (b) allow for unambiguous interpretation of the data when used in the research.

² This will be used by BBMRI-ERIC for ontologizing the data model, in order to make it “machine readable” (allowing for correct interpretation of the data in automated processing workflows).

³ REQUIRED means the data can’t be entered at all without this item being provided; OPTIONAL means data may or may not be provided, but the item will be ready for inputting the data in as part of the data model; RECOMMENDED is a special subclass of the OPTIONAL, which is highly-recommended to be filled in (intended for items where we need the data but where we know that some sources won’t be able to fill this in and we still want such data not being discarded as invalid).

- The data quality analysis tools shall include the various dimensions of data quality including basic checks for accuracy (e.g., violating ranges, value sets), consistency (e.g., valid time differences - sample processing/storing with regard to sample drawing or interrelated elements - only a female can have menses), completeness (if required items are present and to which extent optional fields are filled).
- Statistical methods for data quality analysis shall be supported (e.g., looking for outliers).
- The data validation process⁴ must be transparent and well documented.
- Presentation of results may be provided using comprehensible form including graphical visualizations.

Interoperability

Requirements.

- The API between the local component at the data source and the platform shall be made openly available and documented.
- API to interface to the platform with external services shall be made openly available and documented.
- A reference code example shall be provided that demonstrates the receiving/pull down of a query and reference code to demonstrate the posting of a response.
- A reference code example shall be provided that can demonstrate a user submitting a query to the platform.
- The platform shall support implementation of additional APIs requested by BBMRI-ERIC; the implementation may be either fee-based or contributed to BBMRI-ERIC.
- The platform shall be capable of interfacing to OpenID Connect based Authentication and Authorization Service operated by BBMRI-ERIC for authentication and authorization of users.

Optional/bonus features.

- Implementation of BBMRI-ERIC Directory API to update aggregated descriptors of the data source will be awarded bonus.
- Implementation of BBMRI-ERIC Negotiator API to request release of data and/or biological material by the data source will be awarded bonus.
- Implementation of the following APIs will be awarded bonus: GA4GH Beacon, XNAT.
- Support for ingest of data in HL7 FHIR, OMOP, MIABIS Sample/Donor model, DICOM for radiology imaging, OpenSlide-compatible whole slide imaging, data formats specified in GA4GH Genomic Data Toolkit will be awarded bonus.
- Output of aggregate data in general MIABIS CORE data model will be awarded bonus.

⁴ https://en.wikipedia.org/wiki/Data_validation

Data analysis, pooling and data release support

Requirements.

- The platform shall support at least one of the following: a) federated data analysis capability, b) data pooling base analysis capability.
- If federated data analysis capability is implemented, it shall support
 - review of the analysis code by the data sources prior to running it
 - running custom codes designed by BBMRI-ERIC (subject to review by data source)
 - implementation of custom codes in at least one Turing-complete programming language (e.g., R, Python, JavaScript, Java, ...)
- If data pooling is implemented, it has to support
 - pooling pseudonymized data after approval by the data source
 - pooling anonymized data

Optional/bonus features.

- It is a bonus if the platform supports temporary or permanent data pooling functionality in order to expedite access to pooled data for research.

Donor empowerment

Requirements.

- The platform shall support structured information retrieved by the data source from donors consent and make this queryable.

Optional/bonus features.

- Ability to support dynamic consent⁵ will be awarded bonus - by implementing support for dynamic consent in the platform, and by interfacing to existing consent management systems at the data source.
- Ability to provide patient collected data⁶ will be awarded a bonus.
- Ability to interface to commercial services providing consented patient data (e.g., fitness and health apps) will be awarded a bonus.
- Ability of patients to review and correct their data will be awarded a bonus.

⁵ Dynamic consent means that the patient is able to approve/reject projects for which the querying and/or data release request is being done.

⁶ Patient collected data means data which is collected by the patients directly (e.g., monitoring her mood) or indirectly using some tools such as smart wearables.

Requirements on the Offer

As a part of the offer, the document shall be provided describing roles and responsibilities of the node, BBMRI-ERIC and the federated platform provider in relation to maintenance, installation and other procedures relating to the software, infrastructure and security maintenance, such as backups, security patching and transition plans for migration when operating systems reach end of life.