ADOPT BBMRI-ERIC
GRANT AGREEMENT NO 676550

DELIVERABLE REPORT

Deliverable no	D3.1
Deliverable Title	Consolidated Registry of BBMRI-ERIC Biobanks
Contractual delivery month	M12 (September 2016)
Responsible Partner	BBMRI-ERIC
Author(s)	Petr Holub, David van Enckevort, Robert Reihs, Morris Swertz, Heimo Müller, Araceli DiezFraile, Michael Hummel, Jan-Eric Litton

CONSOLIDATED REGISTRY OF BBMRI-ERIC BIOBANKS
Executive Summary

Biobanks have become the fundamental resource for large scale epidemiology and medical research, in order to develop precision/personalized medicine and optimize public health strategies. Yet, still it is enormously difficult and time intensive to find relevant samples across the thousands of biobanks, due to incompatible sample and data description semantics, data fragmentation and gaps, and heterogeneity in quality and regulations. In this paper we hypothesize that the biobanking community, united in BBMRI-ERIC, can learn from distributed data infrastructure in computer science to deliver a scalable distributed inventory for bioresources, to achieve truly FAIR samples and data: findable, accessible, interoperable and reusable. In this Deliverable we demonstrate implementation of BBMRI-ERIC Directory, where various national biobank inventories have been integrated using symmetric and easy to implement IT protocols that allow data to flow back and forth between the central biobank organizations and the individual biobanks/networks. All software and interfaces are open-source and we invite the wider community to join the BBMRI-ERIC Directory network of biobanks. ADOPT BBMRI-ERIC allowed implementing Directory 2.0, which delivered major improvements over the previous version: (i) completely redesigned hierarchical data model to make it fully compliant with the MIABIS 2.0 Core, which has been finalized just before start of ADOPT BBMRI-ERIC (previous Directory 1.0 used very simple flat data model of biobanks entities only), (ii) implementation of quality checks, (iii) updated web-based user interface to support the new data model. The main part of this Deliverable describes the state-of-implementation of the whole Directory 2.0, as this is most readers, but more detailed discussion of the changes enabled by ADOPT BBMRI-ERIC is in the Section 8.

COPYRIGHT NOTICE

DOCUMENT LOG

Issue	Date	Comment	Author
D3.1rev1	2017-08-18	Update of the Executive Summary and Section 4 (UTF-8 strings) based on reviewers comments. Update of of EU funding recognition to comply with Grant Agreement art. 29.4.	Petr Holub

Contents

1 Introduction

Biobanks[1] are well-organized repositories of biological material that have become an indispensable resource to better understand the epidemiology and biological mechanisms of disease and the fundamental resource for advancing medical research. Many of these collections have now grown to include data from over 100,000s of individuals,[2] but still many research questions require samples and data from multiple collections to reach sufficient statistical power or to achieve sufficient numbers of subjects having particular characteristics; while this has been mostly the domain of rare diseases in the past, with the ongoing development toward more personalized precision medicine the detailed stratification of patients results in similar problems with finding sufficient set of research participants with the same characteristics. Yet biobanks still face large challenges to become more findable and accessible by users on the national and global scales. These challenges range from fragmentation of data structure, lack of availability of data,[3–5] lack of consistent quality management and traceability,[6–9] to fragmentation of privacy protection regulations[10–14] and challenges of scalable secure storage and processing of privacy-sensitive big data.[15–17]

In this paper we address these challenges by providing a scalable inventory on these bioresources called BBMRI-ERIC Directory, describing availability of various resource types such as biological material, data types, expertise, and offered services, as a basis sample/data access interactions between the biobanks as resource/service providers and their users/collaborators. There have been various terms used for this type of services, including sometimes generic overloaded words "catalogs" and "registries". In general, these systems cover various types of information that cannot be considered privacy-sensitive and thus can be shared in a open access mode, with limitations only imposed by the given bioresource business model and not privacy concerns.

For resources to be findable and usable, it is essential that samples and data are described using comparable (harmonized) semantics, so that it is possible to implement efficient search and filtering. There have been a number of attempts to improve the situation with availability and consistency of the inventory data about the infrastructures themselves in the last decade both internationally and nationally. Prominent international examples include P[3]G Observatory,[18] BBMRI Preparatory Phase Catalogue,[4] ISBER Resource Locator,[a] Maelstrom Repository,[b] BBMRI-LPC catalogs,[c] or RD-CONNECT Catalogue[d] on rare diseases.[19] Within more localized scope, some BBMRI-ERIC National Nodes have built their own catalogs which required manual insertion of data, such as Deutsches Biobanken-Register[e] by TMF. While being very valuable for helping to organize biobanking and bioresources communities, these tools also demonstrate the *key deficiency of such centrally-built and managed systems: because of the lack of automated data updates, the information becomes sooner or later obsolete and thus of limited use* for the users. Moreover, users interested in large collections or rare samples still need to utilize multiple inventories at the same time to find relevant materials, as none of the previous ones seemed to have ambition to integrate other inventories. Currently the BBMRI-ERIC Directory is Europe-wide in terms of its content but architecturally and technologically it is ready to scale up to global level and it is capable of aggregating data from various sources.

---

[a] https://www.irlocator.isber.org/ and information on IRL Working Group at http://www.isber.org/page/IRL

[b] https://www.maelstrom-research.org/repository

[c] http://www.bbmri-lpc-biobanks.eu/catalogue.html
  http://bbmri-lpc.iarc.fr/mica/?q=variable-search

[d] http://catalogue.rd-connect.eu/

[e] http://www.biobanken.de/

Meanwhile, successful distributed information systems are well-known in the domain of distributed computer infrastructures, such as distributed grid computing systems,[20] with various architectures have been explored, ranging from client-server communication schemes[21,22] to peer-to-peer systems.[23–25] The biobanking community and bioresources communities in general have to learn from these endeavors and take a similar approach with *(a)* distributed architecture that allows for information flow from the original sources to the inventory services (and back), using *(b)* well-defined stable application programming interfaces (APIs) that allow for their implementation in the biobank information management systems, *(c)* clear component-based architecture that allows for simple implementation of relevant data extraction and processing components as close to the original information sources as possible, while also *(d)* allowing for efficient aggregation of the data to avoid overloading of the infrastructure. With such a structure, the IT systems can behave as a multi-cellular organism, not just as isolated islands often with more or less obsolete information. From the organizational perspective, this needs to be accompanied by a long-term infrastructural commitment of the biobanks as well as their funding organizations, otherwise the initial investment into connecting to such infrastructure may not be justifiable, albeit the costs may be relatively low.

2 Methods

2.1 Use Cases

In order to address these challenges, BBMRI-ERIC has started to develop BBMRI-ERIC Directory as a its first tool, with the following use cases in mind:

[UC-1] *Finding samples and/or data by biomedical and bioinformatics researchers..*

Most common search criteria include material type and diagnosis.[26] This use case involves various accompanying data (e.g., clinical information) and data generated from samples (e.g., genomic data).

[UC-2] *Finding services to host samples/data by biomedical and bioinformatics researchers.*

This includes search for various services offered by the biobanks: sample and data hosting and long term −80°C storage, laboratory facilities for sample processing, expertise in molecular data generation from samples and their interpretation.

[UC-3] *Industrial researchers looking for biobanks with sufficient standards of operations for providing services for industry or to implement joint research.*

Industrial researchers have often more stringent requirements related to intellectual property rights protection and following particular standard operating procedures (SOPs) as a part of quality management. On the other hand, many academic and non-for-profit biobanks have self-imposed or externally imposed restrictions on the collaboration with commercial companies.

[UC-4] *Participants (donors/patients) and their organizations interested to see where their samples might be used and for what purposes.*

Trust of research participants is very important for development of biobanks and being able to demonstrate how the samples and data are used is one of cornerstones of building trustworthy infrastructures. The insights are, however, limited by the requirements on privacy protection.

[UC-5] *Biobank operators to figure out similar biobanks (experience sharing, collaboration, etc.) and to promote their visibility.*

Biobanking is an expensive endeavor and biobank operators can greatly benefit from economies of scope and scale when collaborating. In addition, research requires increasingly large collections for which pooling of data across biobanks is of great added value and biobanks need to support this process consistently.

[UC-6] *Policy makers and funding bodies looking into the extent and use statistics of funded infrastructures.*

Many biobanks are publicly funded and are perceived as very valuable for society. To underline this importance funders and policy makers are looking for data to justify their further investments.

2.2 Data Model

To enable samples and data to be searched in a comparable way, the first development step was designing an extensible data model, that covers all three key components of biobanks: (a) *biological material and associated physical storage facilities*, (b) *data and associated data storage facilities*, and (c) *expertise of the biobankers*.

The core of the data model for the Directory 2.0 relies on to MIABIS 2.0,[27] a standard data model for biobanking, which is evolution of the previously published MIABIS model.[28] As shown in Figure 1, this includes the following basic entities:

- **biobanks** are the institutional units hosting collections of samples and data, as well as providing expertise and other services to their users. This entity does not contain directly any attributes related to the samples or data, which are implemented via links to the collections that are available in the given biobank.

- **collections** are containers for sample sets and/or data sets, with support for recursive creation of sub-collections (of arbitrary finite depth); here properties of the samples and data can be described in aggregated form such as sample counts, diseases, material types, data types, gender, etc.;

- **networks of biobanks** (not defined in the MIABIS 2.0), which may include either whole biobanks or even individual collections inside the biobanks;

- auxiliary **contact information** contact information attached to biobanks, collections and networks needed to get access to samples or data (which is defined centrally to minimize redundancy in the information model).

The data model has been defined in a modular way such that auxiliary classes can be added to suit the needs of biobank (sub)communities, such as to describe clinical, population, research study based, non-human, and standalone collections. Particularly clinical collections are used to enforce existence of attributes describing available diagnoses (which is optional for other types), as it is among the most common search criteria.[26] Standalone collections are used in the countries with legal requirements on institutionalized biobanks, if there are some collections that do not meet these requirements (yet).
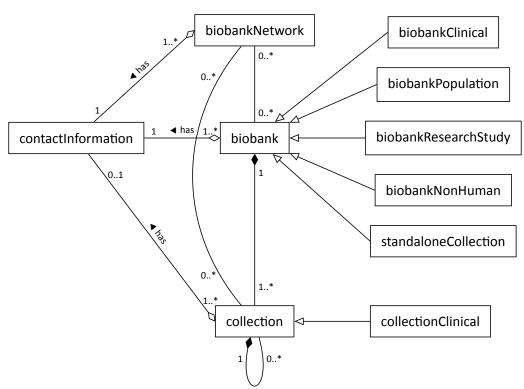
Figure 1: UML class diagram of the BBMRI-ERIC Directory Data Model.
 Note that specialized biobank classes biobankClinical, biboankPopulation, biobankResearch-Study, and biobankNonHuman come as a legacy of collection-less Directory 1.0 and need to be reconsidered in the future development of the model, since these has become properties have been moved to the level of collections in MIABIS 2.0.

Collections. To enable deterministic counts for samples we followed recommendation of MIABIS 2.0[27] that (sub)collections are strictly based on the concept of *set partitioning*: for any collection containing countable (discrete) elements (such as samples/aliquots, images), each element must be exactly in one collection (partition) on the given level of recursion, and there must be no empty collections. This allows for straightforward aggregation: content of each parent entity, be it a collection or a biobank, is a sum of child entities – collections, sub-collections, etc.

The two main groups of material in the collections are *physical material (samples)* and *data*. While samples without data rarely make any sense, the opposite situation with biobanks storing only data is common in many fields, such as imaging biobanks in radiology. In practice, collections are most often created based on (a) purpose of collected material, or (b) life-cycle of the collected material, or (c) funding sources supporting the biobank.

Attributes of entities. Each of the entities have several sets of attributes, as detailed in Section 4: (a) mandatory vs. optional parameters, (b) publicly visible parameters vs. parameters restricted for internal use by BBMRI-ERIC and its National Nodes.

Overall, the attributes can be summarized as follows:

- The attributes for biobanks focus on describing institutional aspects of biobanks, and are anticipated to be extended for other attributes such as available expertise and provided services. The attributes can be grouped into: *(a)* biobank ID (with possible support for mapping of various types of identifiers in the future), *(b)* type of the biobank,[4] *(c)* contact information (via link to contactInformation object with assigned contactPriority, as well as URL), *(d)* head of the biobank and its institutional affiliation, *(e)* information about available information systems (restricted to internal purposes of BBMRI-ERIC and its National Nodes), *(f)* collaboration types supported.

- Attributes describing collections can be divided into:

  - organizational attributes: *(a)* collection ID (with possible support for mapping of various types of identifiers in the future), *(b)* contact information (via link to contactInformation object with assigned contactPriority, as well as URL), *(c)* head of the collection, *(d)* sample and data access policies.

  - attributes describing available physical material and its storage: *(a)* high-level view of stored material types (DNA, plasma, serum, urine, saliva, feces, RNA, blood, frozen tissue or equivalent, FFPE tissue or equivalent, immortalized cell lines), *(b)* type of collection (case control, cohort, cross sectional, longitudinal, twin study, quality control, population based, disease specific, birth cohort, other), *(c)* size of the collection (mandatory 10^n order of magnitude of collected discrete elements – typically samples,[f] with optional exact size with time stamp), *(d)* storage temperatures (based on SPREC 2.0 standard[g]).

  - attributes describing available data: *(a)* available data types (genealogical records, physiological/biochemical measurements, survey data, imaging, medical records), *(b)* access to other data sources (e.g., national registries).

  - attributes describing research participants: *(a)* sex and age of participants, *(b)* available diagnoses (with support for ? and * wildcard characters replacing exactly one and zero or more characters respectively, to allow specification of whole classes of diagnoses, with appropriate search functionality in the Directory user interfaces).

- Biobank networks use attributes describing their institutional aspects as well as commonalities shared by the biobanks/collections participating in the given network.

  Participation of biobanks and collections in biobank networks is implemented via reference attributes from the biobanks and collections. Note *m:n* mapping between biobanks/collections and biobank networks, as one biobank/collection can participate in several biobank networks, and vice versa, each biobank network typically has more than one biobank/collection.

  The attributes can be grouped into: *(a)* biobank network ID (with possible support for mapping of various types of identifiers in the future), *(b)* contact information (via link to contactInformation object with assigned contactPriority, as well as URL), *(c)* commonalities of biobanks participating in the network (collection focus, charter, SOPs, data and sample access policies, MTA/DTA, URL,

---

[f] We consider the search for exact number of samples meaningless before there is consensus on *sample and aliquot definition*, or having these terms standardized possibly as a part of ISO TC 276. We would also advise against abandoning these terms and using number of participants, as has already happened in some Nordic population biobanks, since such approach does not allow to differentiate between a biobank that collects one sample per participant and a time-consistent series of samples per each participant.

[g] `http://www.isber.org/?page=SPREC`

or even complete representation where participating biobanks are only reachable via biobank network), *(d)* head of the biobank network and its institutional affiliation.

Properties of a flat data model.   Similar to the other aggregate biobank inventory systems mentioned above, the Directory also uses a flat data model, inherent to systems not having access to the level of individual samples or data sets about individual persons. This can be perceived as a trade-off between simplicity of implementation (including issues related to protection of information that can be considered personal) and acceptable level of semantic capabilities. The flat data model *prevents a user from asking true "AND questions":* the user can search for the biobanks that have samples/data for certain diagnosis and certain material type, but there is no guarantee there is a combination of these two, i.e., material type for the given diagnosis. Therefore the search using this data model can be understood as filtering out biobanks that certainly do *not* have samples or data relevant for the given search criteria. The result of the search is a set of candidate biobanks, i.e., *biobanks that potentially might have samples for the given purpose*; it is necessary to follow up with these biobanks using per-sample search services of individual biobanks or direct communication (or via planned BBMRI-ERIC Sample Locator and Negotiator systems in the future).

History of the data model.   Directory 1.0 was released in July 2015 with basic support for biobank entities in the data model with contact information embedded for each biobank, particularly because at the time of implementation, revisions of the MIABIS 2.0 Core[27] were still ongoing and the semantics and properties of collections were not clear. The data model has been extended to full MIABIS 2.0 Core compliance in Directory 2.0 in December 2015, supporting biobanks, collections, and biobank networks. These first two versions aim at the following use cases: [UC-1], [UC-3], [UC-5], and [UC-6] (partially, no resource usage information yet).

2.3 Architecture

We early on learned that previously centralized solutions did not scale. So the Directory must operate as a truly distributed network for biobank data aggregation, where new sources and intermediate nodes can be quickly added without central coordination. Conceptually, we followed the proven distributed model of directory services, such as implemented by LDAP (Lightweight Directory Access Protocol[29]): multiple biobanks can connect to a shared directory instance, for example all Dutch biobanks can aggregate to the BBMRI.nl national biobank directory.[h] Then again, Directory instances can connect to a more aggregated Directory instance, for example the BBMRI.nl directory is aggregated into BBMRI-ERIC Directory. This model allows for multiple layers without losing locality on who is in charge of maintaining the data so consistency is ensured. For example, some biobanks may want to first aggregate into an institute/university directory before moving national/European, or for example in Rare disease there are already networks like RD-Connect that can be added as separate directory Node.

We also learned that to motivate local biobank catalogue owners, networks or national nodes to share their data they should be immediate benefit. This lead to the core idea that data sharing must be symmetric form the start, i.e., existing biobank inventories that share data into the Directory network can also immediately access data shared by others.

---

[h] `http://catalogue.bbmri.nl`

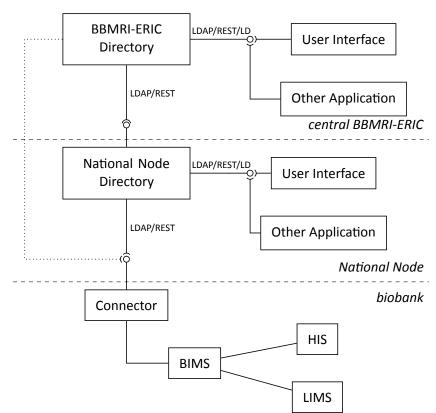[i] `https://developers.google.com/maps/documentation/geocoding/`

Figure 2: BBMRI-ERIC Directory architecture.
Legend: BIMS … biobank information management system, LIMS … laboratory information management system, HIS … hospital information system.

Figure 2 summarizes the resulting distributed system having a multi-layer configuration, with three layers being the default for BBMRI-ERIC member countries: *(a)* biobanks, *(b)* BBMRI-ERIC National Nodes (countries), *(c)* central BBMRI-ERIC infrastructure. To proof that interoperability concept can be easily implemented we created two implementations using LDAP with a LifeRay user interface and using MOLGENIS scientific data platform.[30] In addition we implemented connections to automatically load data from national nodes. Currently we have three data exchange flavors available relying on LDAP[29] or REST/JSON[31] interface for Directory 1.0/2.0, but supporting batch data sharing in tabular formats (e.g., less technical networks can provide collection descriptions via automatic upload of Microsoft Excel and CSV files).

The National Node level and the central level are shown together with user interfaces in the Figure 2. The number of layers can be dynamically adjusted: National Nodes with internal hierarchical structure may introduce additional layers, while may even choose the central layer of the Directory to host the data directly or install a Directory within a hospital for local Directory services.

2.4 Data Quality Checks

In order to maximize usability of the Directory for its users, a data quality check tool has been implemented with support for the following types of checks in LDAP:
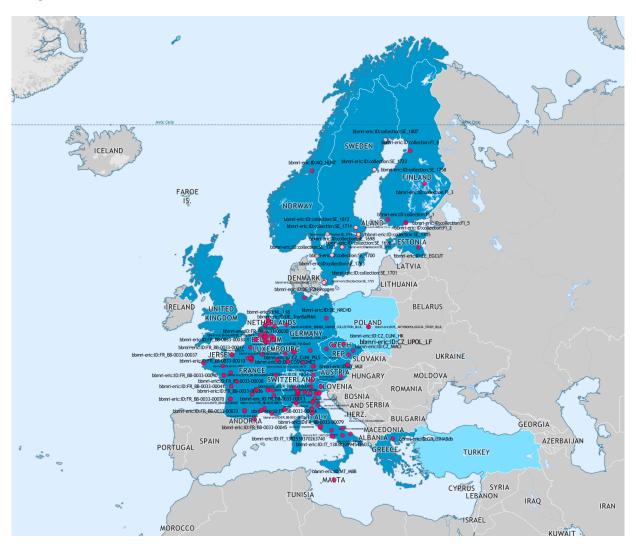
Figure 3: Map visualization of biobanks BBMRI-ERIC Directory the either provided their geographical co-ordinates or that provided an address resolvable via Google Geocoding API.[j]

- *Identifier consistency checks* for checks that are beyond the capability of LDAP schema, including checks for proper format of identifiers and existence of identifiers cross-referenced by entities (contactIDRef, biobankNetworkIDRef).

- *Data syntax checks* for data structures beyond the capability of LDAP schema, including *(a)* compliance of phone numbers to the prescribed format based on E.164 standard and compliance of format of email addresses to RFC 822, *(b)* compliance of URL *(c)* compliance of advertised available diagnoses to the specification (`urn:` prefix followed by ontology specification, e.g., `urn:miriam:icd:` for ICD-10[j]).

---

[j] Other ontologies are supported in the Directory 1.0/2.0. Particularly SNOMED CT is formally supported, but not used in practice because of problems related to transitive licensing requirements of SNOMED CT and lack of support for semantic translation, planned for future releases of the Directory.

Figure 4: LifeRay and MOLGENIS based user interfaces of Directory 2.0/1.0 (MOLGENIS is not yet implemented for 2.0).

- *Data semantics checks* including *(a)* test of reachability of URLs and *(b)* test of reachability of email addresses.[k]

- *Suspicious data checks* including checking *(a)* similarity of data in the directory to the example data (provided to national nodes to simplify initial adoption), *(b)* suspiciously broad specification of available diagnoses (such as advertising availability of all the diagnoses `urn:miriam:icd:*` – which may be however still an acceptable initial approximation for large clinical biobanks), *(c)* sample collections without any material types available and use of deprecated material types (coming from pre-MIABIS 2.0 data structure in Directory 1.0), *(d)* occurrence of suspicious strings in fields (such as "N/A" in mandatory fields).

- *Checks for missing recommended data*, which is implemented as optional data in the LDAP schema, such as biobank description and acronym.

Results of the data checks are provided on regular basis to the BBMRI-ERIC National Node directors (as guarantees of data) or directly to data contributors (if outside of National Node responsibility) for their consideration.

3 Results

Researchers, biobankers and other users can now access the top level Directory within the website of BBMRI-ERIC,[l] see Figure 4. However, the modular component architecture allows also for other user interfaces as well. To demonstrate that this indeed works, we have connected multiple national biobank directories and also various separate software implementations of data model and programmatic interfaces. This functionality has been demonstrated for the Directory 1.0 release in July 2015, with the BBMRI.nl

---

[k] This test is based on Mail::CheckUser Perl module, which attempts direct contact to SMTP servers and uses combination of MAIL and RCPT commands to check for existence of the mailbox. This test is not always reliable as some servers confirm non-existent mailboxes (false positives), but it is still a valuable test to point out some non-existent mailboxes without actually sending unsolicited emails to addresses advertised in the Directory.

[l] `http://bbmri-eric.eu/bbmri-eric-directory`

implementing an interface to the global BBMRI-ERIC data integrated into their national BBMRI.nl Catalogue[m] using MOLGENIS scientific data platform.[30] This software is also available for interested readers to download as open source to start their own, local Directory instance.[n]

At the time of release in December 2015, the Directory 2.0 included 515 biobanks and standalone collections, with estimated number of samples exceeding 60,000,000.[o] This covers 136 clinical or disease-specific biobanks and 189 population biobanks, based on the classification proposed in the paper from BBMRI Preparatory Phase.[4] Other biobanks are of mixed type, they might have opted not to provide their classification, or they may be of another specialized type, such as 2 veterinary biobanks (used for human medicine research purposes) or a biobank specializing on fetal samples.

Directory user interface on the BBMRI-ERIC web pages has generated 505 unique page views a month and 961 page views a month on average during August 1 (a week after the release of Directory 1.0 to let the statistics stabilize) to December 31, 2015.

4 Detailed Description of LDAP Object Classes and Attributes

This section provides detailed overview of LDAP object classed defined in the BBMRI-ERIC Directory LDAP schema, together with description of their attributes. The tables with attribute description use the following shorthand notation:

Type Data type, where mapping to LDAP OID types is as follows:

OID	Name	Note
1.3.6.1.4.1.1466.115.121.1.15	string	UTF-8 encoding of ISO-10646 (Unicode), case-insensitive substring search applied
1.3.6.1.4.1.1466.115.121.1.7	boolean	
1.3.6.1.4.1.1466.115.121.1.27	integer	
1.3.6.1.4.1.1466.115.121.1.50	phone	phone number
1.3.6.1.4.1.1466.115.121.1.11	country	two letter country code

C Cardinality, meaning how many times the attribute may be present. LDAP supports single-value and multi-value attributes (giving the upper limit on cardinality to 1 or *n* respectively), which may be further combined mandatory and optional status (giving the lower limit on cardinality to 1 or 0 respectively).

V Visibility, which can be *(i)* P … public, *(ii)* R … restricted to BBMRI-ERIC internal purposes.

4.1 contactInformation

---

[m] http://directory-molgenis.bbmri-eric.eu/

[n] http://molgenis.github.io

[o] Estimate based on mandatory 10^n order of magnitude information in Directory 2.0. We consider the search for exact number of samples meaningless before there is a common consensus on *sample and aliquot definition*, or having these terms standardized possibly via ISO TC 276. We would also advise against abandoning these terms and using number of participants, as has already happened in some Nordic population biobanks, since such approach does not allow to differentiate between a biobank that collects one sample per participant and a time-consistent series of samples per each participant.

Attribute	Type	C	V	Description
Mandatory				
contactID	string	1	P	Contact identifier.
contactEmail	string	1..n	P	Email according to MIABIS 2.0 – MIABIS-2.0-07-D.
contactCountry	country	1	P	Country according to MIABIS 2.0 – MIABIS-2.0-07-H.
Optional				
contactFirstName	string	0..1	P	First name according to MIABIS 2.0 – MIABIS-2.0-07-A.
contactLastName	string	0..1	P	Last name according to MIABIS 2.0 – MIABIS-2.0-07-B.
contactPhone	phone	0..n	P	Phone number according to MIABIS 2.0 including international prefix (+99999999999 form with no spaces) compliant also to E.123 norm – MIABIS-2.0-07-C.
contactAddress	string	0..n	P	Address according to MIABIS 2.0 – MIABIS-2.0-07-E.
contactZIP	string	0..1	P	ZIP according to MIABIS 2.0 – MIABIS-2.0-07-F.
contactCity	string	0..1	P	City according to MIABIS 2.0 – MIABIS-2.0-07-G.

4.2 collaborationStatus

Attribute	Type	C	V	Description
Optional				
collaborationPartners-Commercial	boolean	0..1	P	Biobank/collection can be used for collaboration with commercial partners.
collaborationPartners-Nonforprofit	boolean	0..1	P	Biobank/collection can be used for collaboration with non-for-profit partners.

4.3 biobank

Description of attributes also includes attributes of the superior objectClasses:

- collaborationStatus

Attribute	Type	C	V	Description
Mandatory				

Continued on next page…

Attribute	Type	C	V	Description
contactIDRef	string	1..n	P	Reference to a contact ID.
contactPriority	integer	1	P	Priority of the contact 1..n (i.e., non-negative integer), where the highest priority should be used for contacting about given set of samples. E.g., if a collection has contactPriority=3, the biobank in which the collection resides has contactPriority=10, and the biobankNetwork to which the collection or biobank belongs has contactPriority=7, the biobank contact should be used.
biobankID	string	1	P	Unique biobank ID withing BBMRI-ERIC based on MIABIS 2.0 standard (ISO 3166-1 alpha-2 + underscore + biobank national ID or name), prefixed with bbmri-eric:ID: string – MIABIS-2.0-01.
biobankName	string	1	P	Biobank name according to MIABIS 2.0 – MIABIS-2.0-03.
biobankJuridicalPerson	string	1..n	P	Juristic person of a biobank according to MIABIS 2.0 – MIABIS-2.0-05.
biobankCountry	country	1..n	P	Country hosting the biobank according to MIABIS 2.0 – MIABIS-2.0-06.
biobankPartnerCharter-Signed	boolean	1	P	Biobank has signed BBMRI-ERIC Partner Charter.
Optional				
bioresourceReference	string	0..n	P	Bioresource reference to be cited when the bioresource (biobank/collection) is used for research.
biobankNetworkIDRef	string	0..n	P	Reference to a biobank network ID, to which the collection or biobank belongs; this attribute can also be used for biobank network, where it refers to the superior biobank network).
geoLatitude	string	0..1	P	Latitude of the biobank in the WGS84 system (the one used by GPS), positive is northern hemisphere.
geoLongitude	string	0..1	P	Longitude of the biobank in the WGS84 system (the one used by GPS), positive is to the East of Greenwich.

Attribute	Type	C	V	Description
collaborationPartners-Commercial	boolean	0..1	P	Biobank/collection can be used for collaboration with commercial partners.
collaborationPartners-Nonforprofit	boolean	0..1	P	Biobank/collection can be used for collaboration with non-for-profit partners.
biobankITSupportAvailable	boolean	0..1	R	Is IT support available at the biobank?
biobankITStaffSize	integer	0..1	R	Size of the biobank dedicated IT staff measured as 2^n.
biobankISAvailable	boolean	0..1	R	Has the biobank a computer-based Information System (IS)?
biobankHISAvailable	boolean	0..1	R	Has the biobank on-line or off-line connection to a Hospital Information System (HIS)?
biobankAcronym	string	0..n	P	Biobank acronym – MIABIS-2.0-02.
biobankDescription	string	0..n	P	Biobank description – MIABIS-2.0-08.
biobankURL	string	0..n	P	Biobank URL – MIABIS-2.0-04.
biobankHeadFirstName	string	0..n	P	First name of a person in charge of the biobank.
biobankHeadLastName	string	0..n	P	Last name of a person in charge of the biobank.
biobankHeadRole	string	0..n	P	Official role of the person in charge of the biobank: typically PI or Director.

4.4 biobankClinical

No attributes available for this object class.

4.5 biobankPopulation

No attributes available for this object class.

4.6 biobankResearchStudy

No attributes available for this object class.

4.7 biobankNonHuman

No attributes available for this object class.

4.8 standaloneCollection

No attributes available for this object class.

4.9 collection

Description of attributes also includes attributes of the superior objectClasses:

- collaborationStatus

Attribute	Type	C	V	Description
Mandatory				
collectionID	string	1	P	Unique collection ID withing BBMRI-ERIC based on MIABIS 2.0 standard, constructed from biobankID prefix + :collection: + local collection ID string – MIABIS-2.0-01.
collectionName	string	1	P	Collection name according to MIABIS 2.0 – MIABIS-2.0-03.
materialStoredDNA	boolean	1	P	DNA: collection contains material of this type (MIABIS-2.0-14).
materialStoredPlasma	boolean	1	P	Plasma: collection contains material of this type (MIABIS-2.0-14).
materialStoredSerum	boolean	1	P	Serum: collection contains material of this type (MIABIS-2.0-14).
materialStoredUrine	boolean	1	P	Urine: collection contains material of this type (MIABIS-2.0-14).
materialStoredSaliva	boolean	1	P	Saliva: collection contains material of this type (MIABIS-2.0-14).
materialStoredFaeces	boolean	1	P	Faeces: collection contains material of this type (MIABIS-2.0-14).
materialStoredOther	string	1..n	P	Other: collection contains material of this type (MIABIS-2.0-14).
materialStoredRNA	boolean	1	P	RNA: collection contains material of this type (MIABIS-2.0-14).
materialStoredBlood	boolean	1	P	Blood: collection contains material of this type (MIABIS-2.0-14).
materialStoredTissueFrozen	boolean	1	P	Frozen Tissue without formalin fixation or equivalent: collection contains material of this type (MIABIS-2.0-14).
materialStoredTissueFFPE	boolean	1	P	Tissue, formalin fixated and paraffin preserved or equivalent: collection contains material of this type (MIABIS-2.0-14).

Attribute	Type	C	V	Description
materialStored-ImmortalizedCellLines	boolean	1	P	Immortalized cell lines: collection contains material of this type (MIABIS-2.0-14).
materialStoredIsolated-Pathogen	boolean	1	P	Isolated Pathogen: collection contains material of this type (MIABIS-2.0-14).
collectionTypeCaseControl	boolean	1	P	A case-control study design compares two groups of subjects: those with the disease or condition under study (cases) and a very similar group of subjects who do not have the disease or condition (controls). – EMBL (EFO) – MIABIS-2.0-19.
collectionTypeCohort	boolean	1	P	A form of longitudinal study for the analysis of risk factors following a group of people who do not have a disease, and uses correlations to determine the absolute risk of subject contraction. – Wikipedia (rewritten) – MIABIS-2.0-19.
collectionTypeCross-Sectional	boolean	1	P	A type of observational study that involves data collection from a population, or a representative subset, at one specific point in time. – Wikipedia – MIABIS-2.0-19.
collectionTypeLongitudinal	boolean	1	P	Research studies involving repeated observations of the same entity over time. In the biobank context, longitudinal studies sample a group of people in a given time period, and study them at intervals by the acquisition and analyses of data and/or samples over time. – P3G – MIABIS-2.0-19.
collectionTypeTwinStudy	boolean	1	P	Twin studies measure the contribution of genetics (as opposed to environment) to a given trait or condition of interest. – MIABIS-2.0-19.
collectionTypeQuality-Control	boolean	1	P	A quality control testing study design type is where some aspect of the experiment is quality controlled for the purposes of quality assurance. – EMBL (EFO) – MIABIS-2.0-19.

Attribute	Type	C	V	Description
collectionTypePopulation-Based	boolean	1	P	Study done at the population level or among the population groups, generally to find the cause, incidence or spread of the disease or to see the response to the treatment, nutrition or environment. – Wikipedia (rewritten) – MIABIS-2.0-19.
collectionTypeDisease-Specific	boolean	1	P	A collection for which material and information is collected from subjects that have already developed a particular disease. – EMBL (EFO) – MIABIS-2.0-19.
collectionTypeBirthCohort	boolean	1	P	A corhort study for which the subjects are followed from the time of birth usually including information about gestation and follow up. – MIABIS-2.0-19.
collectionTypeOther	string	1..n	P	Other type of collection text specified (MIABIS-2.0-19).
collectionOrderOf-Magnitude	integer	1	P	Size of the collection measured as 10^n samples.
Optional				
bioresourceReference	string	0..n	P	Bioresource reference to be cited when the bioresource (biobank/collection) is used for research.
contactIDRef	string	0..n	P	Reference to a contact ID.
contactPriority	integer	0..1	P	Priority of the contact 1..n (i.e., non-negative integer), where the highest priority should be used for contacting about given set of samples. E.g., if a collection has contactPriority=3, the biobank in which the collection resides has contactPriority=10, and the biobankNetwork to which the collection or biobank belongs has contactPriority=7, the biobank contact should be used.
biobankNetworkIDRef	string	0..n	P	Reference to a biobank network ID, to which the collection or biobank belongs; this attribute can also be used for biobank network, where it refers to the superior biobank network).
geoLatitude	string	0..1	P	Latitude of the biobank in the WGS84 system (the one used by GPS), positive is northern hemisphere.

Attribute	Type	C	V	Description
geoLongitude	string	0..1	P	Longitude of the biobank in the WGS84 system (the one used by GPS), positive is to the East of Greenwich.
collaborationPartners-Commercial	boolean	0..1	P	Biobank/collection can be used for collaboration with commercial partners.
collaborationPartners-Nonforprofit	boolean	0..1	P	Biobank/collection can be used for collaboration with non-for-profit partners.
collectionAcronym	string	0..1	P	Collection acronym according to MIABIS 2.0 – MIABIS-2.0-02.
collectionDescription	string	0..1	P	Collection description according to MIABIS 2.0 – MIABIS-2.0-08.
collectionSexMale	boolean	0..n	P	The sex of the individuals in the sample collection. – MIABIS-2.0-09.
collectionSexFemale	boolean	0..n	P	The sex of the individuals in the sample collection. – MIABIS-2.0-09.
collectionSexUnknown	boolean	0..n	P	The sex of the individuals in the sample collection. – MIABIS-2.0-09.
collectionSex-Undiferrentiated	boolean	0..n	P	The sex of the individuals in the sample collection. – MIABIS-2.0-09.
collectionAgeLow	integer	0..1	P	Age of youngest sample donor at time of sample donation – MIABIS-2.0-10.
collectionAgeHigh	integer	0..1	P	Age of oldest sample donor at time of sample donation – MIABIS-2.0-11.
collectionAgeUnit	string	0..1	P	Unit defining Age Low and Age High. Can be one of the following values: years, months, weeks, days – MIABIS-2.0-08.
collectionAvailable-BiologicalSamples	boolean	0..1	P	Denotes whether biological samples are available (MIABIS-2.0-13).
collectionAvailableSurvey-Data	boolean	0..1	P	Denotes whether survey data are available (MIABIS-2.0-13).
collectionAvailableImaging-Data	boolean	0..1	P	Denotes whether imaging data are available (MIABIS-2.0-13).
collectionAvailableMedical-Records	boolean	0..1	P	Denotes whether medical records are available (MIABIS-2.0-13).
collectionAvailableNational-Registries	boolean	0..1	P	Denotes whether register data is associated to the participants in the sample collection/study (MIABIS-2.0-13).

Attribute	Type	C	V	Description
collectionAvailable-GenealogicalRecords	boolean	0..1	P	Denotes whether genealogical records are available (MIABIS-2.0-13).
collectionAvailablePhysio-BiochemMeasurements	boolean	0..1	P	Denotes whether physiological/biochemical measurements are available (MIABIS-2.0-13).
collectionAvailableOther	boolean	0..1	P	Denotes whether other samples/data is available (MIABIS-2.0-13).
temperatureRoom	boolean	0..1	P	Sample storage temperature – Room temperature – SPREC 2.0 (MIABIS-2.0-15).
temperature2to10	boolean	0..1	P	Sample storage temperature – between 2 and 10°C – SPREC 2.0 (MIABIS-2.0-15).
temperature-18to-35	boolean	0..1	P	Sample storage temperature – between -18 and -35°C – SPREC 2.0 (MIABIS-2.0-15).
temperature-60to-85	boolean	0..1	P	Sample storage temperature – between -60 and -85°C – SPREC 2.0 (MIABIS-2.0-15).
temperatureLN	boolean	0..1	P	Sample storage temperature – liquid nitrogen, -150 to -196°C (MIABIS-2.0-15).
temperatureOther	string	0..n	P	Sample storage temperature – other, text specified (MIABIS-2.0-15).
diagnosisAvailable	string	0..n	P	Diagnosis available in the collection, with the ontology prefix, possibly using * and ? wildcards, and prefix notation to denote diagnosis nomenclature – so far urn:miriam:icd: prefix for ICD-10, and urn:miriam:snomedct: prefix for SNOMED CT (examples being urn:miriam:icd:C*, urn:miriam:snomedct:25*) – MIABIS-2.0-17, adapted.
collectionHeadFirstName	string	0..n	P	First name of a person in charge of the collection.
collectionHeadLastName	string	0..n	P	Last name of a person in charge of the collection.
collectionHeadRole	string	0..n	P	Official role of the person in charge of the collection: typically PI or Director.
collectionSampleAccessFee	boolean	0..1	P	Denotes whether access to samples may be obtained on fee-based basis.

Attribute	Type	C	V	Description
collectionSampleAccess-JointProjects	boolean	0..1	P	Denotes whether access to samples may be obtained on joint project basis.
collectionSampleAccess-Description	string	0..n	P	Short description of access rules.
collectionDataAccessFee	boolean	0..1	P	Denotes whether access to data may be obtained on fee-based basis.
collectionDataAccessJoint-Projects	boolean	0..1	P	Denotes whether access to data may be obtained on joint project basis.
collectionDataAccess-Description	string	0..n	P	Short description of access rules.
collectionSampleAccessURI	string	0..n	P	URI describing access policy for the samples.
collectionDataAccessURI	string	0..n	P	URI describing access policy for the data.
collectionSize	integer	0..1	P	Exact size of the collection to the given date.
collectionSizeTimestamp	integer	0..1	P	Date to which the size of the collection was valid, absolute time in ISO 8601 format.

4.10 collectionClinical

Attribute	Type	C	V	Description
Mandatory				
diagnosisAvailable	string	1..n	P	Diagnosis available in the collection, with the ontology prefix, possibly using * and ? wildcards, and prefix notation to denote diagnosis nomenclature – so far urn:miriam:icd: prefix for ICD-10, and urn:miriam:snomedct: prefix for SNOMED CT (examples being urn:miriam:icd:C*, urn:miriam:snomedct:25*) – MIABIS-2.0-17, adapted.

4.11 biobankNetwork

Attribute	Type	C	V	Description
Mandatory				
contactIDRef	string	1..n	P	Reference to a contact ID.
contactPriority	integer	1	P	Priority of the contact 1..n (i.e., non-negative integer), where the highest priority should be used for contacting about given set of samples. E.g., if a collection has contactPriority=3, the biobank in which the collection resides has contactPriority=10, and the biobankNetwork to which the collection or biobank belongs has contactPriority=7, the biobank contact should be used.
biobankNetworkID	string	1	P	Unique ID of a biobank network within BBMRI-ERIC based on MIABIS 2.0 standard (ISO 3166-1 alpha-2 + underscore + biobank national ID or name), prefixed with bbmri-eric:bbnet:ID: string; if biobank network is on European or higher level, EU_ prefix is to be used instead of country prefix.
biobankNetworkName	string	1	P	Biobank network name.
biobankNetworkCommon-CollectionFocus	boolean	1	P	All the biobanks/collections in the network share the same focus with which the samples are collected (e.g., disease specific). Further details of the focus should be provided in the biobankNetworkDescription attribute.
biobankNetworkCommon-Charter	boolean	1	P	All the biobanks/collections in the network have to have a network charter signed.
biobankNetworkCommon-SOPs	boolean	1	P	All the biobanks/collections in the network share the same SOPs.
biobankNetworkCommon-DataAccessPolicy	boolean	1	P	All the biobanks/collections in the network share the same data access policy.
biobankNetworkCommon-SampleAccessPolicy	boolean	1	P	All the biobanks/collections in the network share the same sample access policy.
biobankNetworkCommon-MTA	boolean	1	P	All the biobanks/collections in the network share the same MTA.
biobankNetworkCommon-Representation	boolean	1	P	All the biobanks/collections in the network are represented using the network only.
biobankNetworkCommon-URL	boolean	1	P	All the biobanks/collections in the network share the same web presentation on the common URL.

Continued on next page…

Attribute	Type	C	V	Description
Optional				
biobankNetworkIDRef	string	0..n	P	Reference to a biobank network ID, to which the collection or biobank belongs; this attribute can also be used for biobank network, where it refers to the superior biobank network).
geoLatitude	string	0..1	P	Latitude of the biobank in the WGS84 system (the one used by GPS), positive is northern hemisphere.
geoLongitude	string	0..1	P	Longitude of the biobank in the WGS84 system (the one used by GPS), positive is to the East of Greenwich.
biobankNetworkAcronym	string	0..1	P	Biobank network acronym.
biobankNetwork-Description	string	0..1	P	Biobank network description.
biobankNetworkURL	string	0..1	P	Biobank network URL.
biobankNetworkJuridical-Person	string	0..n	P	Juristic person of a biobank network according

5 Listing of biobank.schema

```
1   # BBMRI-ERIC Directory 2.1 Schema
    # $Id: biobank.schema 187 2016-03-14 10:07:00Z hopet $
3   #
    # 1.3.6.1.4.1.45510
5   #                                          .1 = biobank/standalone collection
    #                                          .1.1 = attributes
7   #                                          .1.2 = objectclass
    #                                          .1.3 = syntax (probably never used)
9   #

11  objectIdentifier BBMRIERICDirectoryRoot 1.3.6.1.4.1.45510
    objectIdentifier BBMRIERICDirectoryCommon BBMRIERICDirectoryRoot:1
13  objectIdentifier BBMRIERICDirectoryCommonAttr BBMRIERICDirectoryCommon:1
    objectIdentifier BBMRIERICDirectoryCommonObj BBMRIERICDirectoryCommon:2
15  objectIdentifier BBMRIERICDirectoryBB BBMRIERICDirectoryRoot:2
    objectIdentifier BBMRIERICDirectoryBBAttr BBMRIERICDirectoryBB:1
17  objectIdentifier BBMRIERICDirectoryBBObj BBMRIERICDirectoryBB:2
    objectIdentifier BBMRIERICDirectoryColl BBMRIERICDirectoryRoot:3
19  objectIdentifier BBMRIERICDirectoryCollAttr BBMRIERICDirectoryColl:1
    objectIdentifier BBMRIERICDirectoryCollObj BBMRIERICDirectoryColl:2
21  objectIdentifier BBMRIERICDirectoryBBNet BBMRIERICDirectoryRoot:4
    objectIdentifier BBMRIERICDirectoryBBNetAttr BBMRIERICDirectoryBBNet:1
23  objectIdentifier BBMRIERICDirectoryBBNetObj BBMRIERICDirectoryBBNet:2

25
    #
27  # Bioresource Reference
    #
29
    attributetype ( BBMRIERICDirectoryCommonAttr:1
31   NAME 'bioresourceReference'
     DESC 'Bioresource reference to be cited when the bioresource (biobank/collection) is used for
         research.'
33   EQUALITY caseIgnoreMatch
     SUBSTR caseIgnoreSubstringsMatch
35   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

37
    #
39  # Contact info
    #
41
    attributetype ( BBMRIERICDirectoryCommonAttr:2.1
43   NAME 'contactID'
     DESC 'Contact identifier.'
45   EQUALITY caseIgnoreMatch
     SUBSTR caseIgnoreSubstringsMatch
47   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

49  attributetype ( BBMRIERICDirectoryCommonAttr:2.2
     NAME 'contactFirstName'
51   DESC 'First name according to MIABIS 2.0 - MIABIS-2.0-07-A.'
     EQUALITY caseIgnoreMatch
53   SUBSTR caseIgnoreSubstringsMatch
     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
55
    attributetype ( BBMRIERICDirectoryCommonAttr:2.3
57   NAME 'contactLastName'
     DESC 'Last name according to MIABIS 2.0 - MIABIS-2.0-07-B.'
59   EQUALITY caseIgnoreMatch
     SUBSTR caseIgnoreSubstringsMatch
61   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

63  attributetype ( BBMRIERICDirectoryCommonAttr:2.4
     NAME 'contactPhone'
65   DESC 'Phone number according to MIABIS 2.0 including international prefix (+99999999999 form
         with no spaces) compliant also to E.123 norm - MIABIS-2.0-07-C.'
     SYNTAX '1.3.6.1.4.1.1466.115.121.1.50')
67
    attributetype ( BBMRIERICDirectoryCommonAttr:2.5
69   NAME 'contactEmail'
     DESC 'Email according to MIABIS 2.0 - MIABIS-2.0-07-D.'
```

```
71   EQUALITY caseIgnoreMatch
     SUBSTR caseIgnoreSubstringsMatch
73   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

75   attributetype ( BBMRIERICDirectoryCommonAttr:2.6
      NAME 'contactAddress'
77    DESC 'Address according to MIABIS 2.0 - MIABIS-2.0-07-E.'
      EQUALITY caseIgnoreMatch
79    SUBSTR caseIgnoreSubstringsMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
81
      attributetype ( BBMRIERICDirectoryCommonAttr:2.7
83    NAME 'contactZIP'
      DESC 'ZIP according to MIABIS 2.0 - MIABIS-2.0-07-F.'
85    EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
87    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

89    attributetype ( BBMRIERICDirectoryCommonAttr:2.8
      NAME 'contactCity'
91    DESC 'City according to MIABIS 2.0 - MIABIS-2.0-07-G.'
      EQUALITY caseIgnoreMatch
93    SUBSTR caseIgnoreSubstringsMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
95
      attributetype ( BBMRIERICDirectoryCommonAttr:2.9
97    NAME 'contactCountry'
      DESC 'Country according to MIABIS 2.0 - MIABIS-2.0-07-H.'
99    SYNTAX '1.3.6.1.4.1.1466.115.121.1.11' SINGLE-VALUE)

101  objectClass ( BBMRIERICDirectoryCommonObj:1
      NAME 'contactInformation'
103   DESC 'Contact Information Object - MIABIS-2.0-07.'
      MUST ( contactID $ contactEmail $ contactCountry )
105   MAY ( contactFirstName $ contactLastName $ contactPhone $ contactAddress $ contactCity $
          contactZIP ) )

107
      #
109  # Contact references
      #
111
     attributetype ( BBMRIERICDirectoryCommonAttr:3.1
113   NAME 'contactIDRef'
      DESC 'Reference to a contact ID.'
115   EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
117   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

119  attributetype ( BBMRIERICDirectoryCommonAttr:3.2
      NAME 'contactPriority'
121   DESC 'Priority of the contact 1..n (i.e., non-negative integer), where the highest priority
          should be used for contacting about given set of samples. E.g., if a collection has
          contactPriority=3, the biobank in which the collection resides has contactPriority=10, and
          the biobankNetwork to which the collection or biobank belongs has contactPriority=7, the
          biobank contact should be used.'
      EQUALITY integerMatch
123   SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

125
      #
127  # References for implementation of mapping between biobankNetworks and
      # (biobanks, collections)
129  #

131  attributetype ( BBMRIERICDirectoryCommonAttr:4
      NAME 'biobankIDRef'
133   DESC 'Reference to a biobank ID.'
      EQUALITY caseIgnoreMatch
135   SUBSTR caseIgnoreSubstringsMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
137
      attributetype ( BBMRIERICDirectoryCommonAttr:5
139   NAME 'collectionIDRef'
      DESC 'Reference to a collection ID.'
```

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 676550.

26/49

```
141   EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
143   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

145   attributetype ( BBMRIERICDirectoryCommonAttr:6
      NAME 'biobankNetworkIDRef'
147   DESC 'Reference to a biobank network ID, to which the collection or biobank belongs; this
          attribute can also be used for biobank network, where it refers to the superior biobank
          network).'
      EQUALITY caseIgnoreMatch
149   SUBSTR caseIgnoreSubstringsMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

151

153   #
      # Georeferencing attributes
155   #

157   attributetype ( BBMRIERICDirectoryCommonAttr:7.1
      NAME 'geoLatitude'
159   DESC 'Latitude of the biobank in the WGS84 system (the one used by GPS), positive is northern
          hemisphere.'
      EQUALITY caseIgnoreMatch
161   SUBSTR caseIgnoreSubstringsMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
163
      attributetype ( BBMRIERICDirectoryCommonAttr:7.2
165   NAME 'geoLongitude'
      DESC 'Longitude of the biobank in the WGS84 system (the one used by GPS), positive is to the
          East of Greenwich.'
167   EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
169   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

171

      #
173   # Collaboration options
      #
175
      attributetype ( BBMRIERICDirectoryCommonAttr:8.1
177   NAME 'collaborationPartnersCommercial'
      DESC 'Biobank/collection can be used for collaboration with commercial partners.'
179   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
181
      attributetype ( BBMRIERICDirectoryCommonAttr:8.2
183   NAME 'collaborationPartnersNonforprofit'
      DESC 'Biobank/collection can be used for collaboration with non-for-profit partners.'
185   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
187
      objectClass ( BBMRIERICDirectoryCommonObj:2
189   NAME 'collaborationStatus'
      DESC 'Abstract class describing collaboration status, to be used for biobanks and collections.'
191   ABSTRACT
      MAY ( collaborationPartnersCommercial $ collaborationPartnersNonforprofit ) )
193


195
      #
197   # Biobank definition
      #
199
      attributetype ( BBMRIERICDirectoryBBAttr:1
201   NAME 'biobankID'
      DESC 'Unique biobank ID withing BBMRI-ERIC based on MIABIS 2.0 standard (ISO 3166-1 alpha-2 +
          underscore + biobank national ID or name), prefixed with bbmri-eric:ID: string -
          MIABIS-2.0-01.'
203   EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
205   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

207   attributetype ( BBMRIERICDirectoryBBAttr:2
      NAME 'biobankName'
209   DESC 'Biobank name according to MIABIS 2.0 - MIABIS-2.0-03.'
```

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 676550.

27/49

```
       EQUALITY caseIgnoreMatch
211    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
213
       attributetype ( BBMRIERICDirectoryBBAttr:3
215    NAME 'biobankJuridicalPerson'
       DESC 'Juristic person of a biobank according to MIABIS 2.0 - MIABIS-2.0-05.'
217    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
219    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

221    attributetype ( BBMRIERICDirectoryBBAttr:4
       NAME 'biobankCountry'
223    DESC 'Country hosting the biobank according to MIABIS 2.0 - MIABIS-2.0-06.'
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.11')
225


227

229    # IT availability

231    attributetype ( BBMRIERICDirectoryBBAttr:5.1
       NAME 'biobankITSupportAvailable'
233    DESC 'Is IT support available at the biobank?'
       EQUALITY booleanMatch
235    SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)

237    attributetype ( BBMRIERICDirectoryBBAttr:5.2
       NAME 'biobankITStaffSize'
239    DESC 'Size of the biobank dedicated IT staff measured as 2^n.'
       EQUALITY integerMatch
241    SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

243    attributetype ( BBMRIERICDirectoryBBAttr:5.3
       NAME 'biobankISAvailable'
245    DESC 'Has the biobank a computer-based Information System (IS)?.'
       EQUALITY booleanMatch
247    SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)

249    attributetype ( BBMRIERICDirectoryBBAttr:5.4
       NAME 'biobankHISAvailable'
251    DESC 'Has the biobank on-line or off-line connection to a Hospital Information System (HIS)?.'
       EQUALITY booleanMatch
253    SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)

255
       # Partner charter
257
       attributetype ( BBMRIERICDirectoryBBAttr:6
259    NAME 'biobankPartnerCharterSigned'
       DESC 'Biobank has signed BBMRI-ERIC Partner Charter.'
261    EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
263
       # Additional biobank-level attributes from MIABIS
265
       attributetype ( BBMRIERICDirectoryBBAttr:7
267    NAME 'biobankAcronym'
       DESC 'Biobank acronym - MIABIS-2.0-02.'
269    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
271    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' )

273    attributetype ( BBMRIERICDirectoryBBAttr:8
       NAME 'biobankDescription'
275    DESC 'Biobank description - MIABIS-2.0-08.'
       EQUALITY caseIgnoreMatch
277    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' )
279
       attributetype ( BBMRIERICDirectoryBBAttr:9
281    NAME 'biobankURL'
       DESC 'Biobank URL - MIABIS-2.0-04.'
283    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
```

```
285   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' )

287   # Biobank head

289   attributetype ( BBMRIERICDirectoryBBAttr:10.1
       NAME 'biobankHeadFirstName'
291    DESC 'First name of a person in charge of the biobank.'
       EQUALITY caseIgnoreMatch
293    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
295
       attributetype ( BBMRIERICDirectoryBBAttr:10.2
297    NAME 'biobankHeadLastName'
       DESC 'Last name of a person in charge of the biobank.'
299    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
301    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

303   attributetype ( BBMRIERICDirectoryBBAttr:10.3
       NAME 'biobankHeadRole'
305    DESC 'Official role of the person in charge of the biobank: typically PI or Director.'
       EQUALITY caseIgnoreMatch
307    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
309
311   # Definition of various biobank types - using object classes allows us to
      # create various combinations of mandatory/optional info
313
      objectClass ( BBMRIERICDirectoryBBObj:1
315    NAME 'biobank'
       DESC 'Biobank (or standalone collection) Object'
317    SUP collaborationStatus
       MUST ( biobankID $ biobankName $ biobankJuridicalPerson $ biobankCountry $
           biobankPartnerCharterSigned $ contactIDRef $ contactPriority )
319    MAY ( biobankITSupportAvailable $ biobankITStaffSize $ biobankISAvailable $ biobankHISAvailable
           $ biobankAcronym $ biobankDescription $ biobankURL $ biobankHeadFirstName $
           biobankHeadLastName $ biobankHeadRole $ bioresourceReference $ biobankNetworkIDRef $
           geoLatitude $ geoLongitude ) )
321   objectClass ( BBMRIERICDirectoryBBObj:2
       NAME 'biobankClinical'
323    DESC 'Clinical biobank Object.'
       AUXILIARY )
325
      objectClass ( BBMRIERICDirectoryBBObj:3
327    NAME 'biobankPopulation'
       DESC 'Population biobank Object'
329    AUXILIARY )

331
      objectClass ( BBMRIERICDirectoryBBObj:4
333    NAME 'biobankResearchStudy'
       DESC 'Research/study-based biobank Object'
335    AUXILIARY )

337
      objectClass ( BBMRIERICDirectoryBBObj:5
339    NAME 'biobankNonHuman'
       DESC 'Non Human biobank Object'
341    AUXILIARY )

343   objectClass ( BBMRIERICDirectoryBBObj:6
       NAME 'standaloneCollection'
345    DESC 'Standalone collection Object standing outside of a biobank'
       AUXILIARY)
347

349   # Adding support for embedded collections

351   attributetype ( BBMRIERICDirectoryCollAttr:1
               NAME 'collectionID'
353    DESC 'Unique collection ID withing BBMRI-ERIC based on MIABIS 2.0 standard, constructed from
           biobankID prefix + :collection: + local collection ID string - MIABIS-2.0-01.'
       EQUALITY caseIgnoreMatch
```

```
355    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
357
       attributetype ( BBMRIERICDirectoryCollAttr:2
359     NAME 'collectionAcronym'
        DESC 'Collection acronym according to MIABIS 2.0 - MIABIS-2.0-02.'
361     EQUALITY caseIgnoreMatch
        SUBSTR caseIgnoreSubstringsMatch
363     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

365    attributetype ( BBMRIERICDirectoryCollAttr:3
        NAME 'collectionName'
367     DESC 'Collection name according to MIABIS 2.0 - MIABIS-2.0-03.'
        EQUALITY caseIgnoreMatch
369     SUBSTR caseIgnoreSubstringsMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
371
       attributetype ( BBMRIERICDirectoryCollAttr:4
373     NAME 'collectionDescription'
        DESC 'Collection description according to MIABIS 2.0 - MIABIS-2.0-08.'
375     EQUALITY caseIgnoreMatch
        SUBSTR caseIgnoreSubstringsMatch
377     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)

379    # Expanded Sex from text list to booleans

381    attributetype ( BBMRIERICDirectoryCollAttr:5.1
        NAME 'collectionSexMale'
383     DESC 'The sex of the individuals in the sample collection. - MIABIS-2.0-09.'
        EQUALITY booleanMatch
385     SYNTAX '1.3.6.1.4.1.1466.115.121.1.7')

387    attributetype ( BBMRIERICDirectoryCollAttr:5.2
        NAME 'collectionSexFemale'
389     DESC 'The sex of the individuals in the sample collection. - MIABIS-2.0-09.'
        EQUALITY booleanMatch
391     SYNTAX '1.3.6.1.4.1.1466.115.121.1.7')

393    attributetype ( BBMRIERICDirectoryCollAttr:5.3
        NAME 'collectionSexUnknown'
395     DESC 'The sex of the individuals in the sample collection. - MIABIS-2.0-09.'
        EQUALITY booleanMatch
397     SYNTAX '1.3.6.1.4.1.1466.115.121.1.7')

399    attributetype ( BBMRIERICDirectoryCollAttr:5.4
        NAME 'collectionSexUndiferrentiated'
401     DESC 'The sex of the individuals in the sample collection. - MIABIS-2.0-09.'
        EQUALITY booleanMatch
403     SYNTAX '1.3.6.1.4.1.1466.115.121.1.7')

405    # Age-related specs

407    attributetype ( BBMRIERICDirectoryCollAttr:6.1
        NAME 'collectionAgeLow'
409     DESC 'Age of youngest sample donor at time of sample donation - MIABIS-2.0-10.'
        EQUALITY integerMatch
411     SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

413    attributetype ( BBMRIERICDirectoryCollAttr:6.2
        NAME 'collectionAgeHigh'
415     DESC 'Age of oldest sample donor at time of sample donation - MIABIS-2.0-11.'
        EQUALITY integerMatch
417     SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

419    attributetype ( BBMRIERICDirectoryCollAttr:6.3
        NAME 'collectionAgeUnit'
421     DESC 'Unit defining Age Low and Age High. Can be one of the following values: years, months,
            weeks, days - MIABIS-2.0-08.'
        EQUALITY caseIgnoreMatch
423     SUBSTR caseIgnoreSubstringsMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
425
       # sample/data types available
427
       attributetype ( BBMRIERICDirectoryCollAttr:7.1
```

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 676550.

30/49

```
429   NAME 'collectionAvailableBiologicalSamples'
      DESC 'Denotes whether biological samples are available (MIABIS-2.0-13).'
431   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
433
      attributetype ( BBMRIERICDirectoryCollAttr:7.2
435   NAME 'collectionAvailableSurveyData'
      DESC 'Denotes whether survey data are available (MIABIS-2.0-13).'
437   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
439
      attributetype ( BBMRIERICDirectoryCollAttr:7.3
441   NAME 'collectionAvailableImagingData'
      DESC 'Denotes whether imaging data are available (MIABIS-2.0-13).'
443   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
445
      attributetype ( BBMRIERICDirectoryCollAttr:7.4
447   NAME 'collectionAvailableMedicalRecords'
      DESC 'Denotes whether medical records are available (MIABIS-2.0-13).'
449   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
451
      attributetype ( BBMRIERICDirectoryCollAttr:7.5
453   NAME 'collectionAvailableNationalRegistries'
      DESC 'Denotes whether register data is associated to the participants in the sample
            collection/study (MIABIS-2.0-13).'
455   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
457
      attributetype ( BBMRIERICDirectoryCollAttr:7.6
459   NAME 'collectionAvailableGenealogicalRecords'
      DESC 'Denotes whether genealogical records are available (MIABIS-2.0-13).'
461   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
463
      attributetype ( BBMRIERICDirectoryCollAttr:7.7
465   NAME 'collectionAvailablePhysioBiochemMeasurements'
      DESC 'Denotes whether physiological/biochemical measurements are available (MIABIS-2.0-13).'
467   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
469
      attributetype ( BBMRIERICDirectoryCollAttr:7.8
471   NAME 'collectionAvailableOther'
      DESC 'Denotes whether other samples/data is available (MIABIS-2.0-13).'
473   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
475
      # Material types stored
477
      attributetype ( BBMRIERICDirectoryCollAttr:8.1
479   NAME 'materialStoredDNA'
      DESC 'DNA: collection contains material of this type (MIABIS-2.0-14).'
481   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
483
      attributetype ( BBMRIERICDirectoryCollAttr:8.2
485   NAME 'materialStoredPlasma'
      DESC 'Plasma: collection contains material of this type (MIABIS-2.0-14).'
487   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
489
      attributetype ( BBMRIERICDirectoryCollAttr:8.3
491   NAME 'materialStoredSerum'
      DESC 'Serum: collection contains material of this type (MIABIS-2.0-14).'
493   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
495
      attributetype ( BBMRIERICDirectoryCollAttr:8.4
497   NAME 'materialStoredUrine'
      DESC 'Urine: collection contains material of this type (MIABIS-2.0-14).'
499   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
501
      attributetype ( BBMRIERICDirectoryCollAttr:8.5
```

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 676550.

31/49

```
503   NAME 'materialStoredSaliva'
      DESC 'Saliva: collection contains material of this type (MIABIS-2.0-14).'
505   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
507
      attributetype ( BBMRIERICDirectoryCollAttr:8.6
509   NAME 'materialStoredFaeces'
      DESC 'Faeces: collection contains material of this type (MIABIS-2.0-14).'
511   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
513
      attributetype ( BBMRIERICDirectoryCollAttr:8.7
515   NAME 'materialStoredOther'
      DESC 'Other: collection contains material of this type (MIABIS-2.0-14).'
517   EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
519   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' )
521   attributetype ( BBMRIERICDirectoryCollAttr:8.8
      NAME 'materialStoredRNA'
523   DESC 'RNA: collection contains material of this type (MIABIS-2.0-14).'
      EQUALITY booleanMatch
525   SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
527   attributetype ( BBMRIERICDirectoryCollAttr:8.9
      NAME 'materialStoredBlood'
529   DESC 'Blood: collection contains material of this type (MIABIS-2.0-14).'
      EQUALITY booleanMatch
531   SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
533   attributetype ( BBMRIERICDirectoryCollAttr:8.10
      NAME 'materialStoredTissueFrozen'
535   DESC 'Frozen Tissue without formalin fixation or equivalent: collection contains material of
          this type (MIABIS-2.0-14).'
      EQUALITY booleanMatch
537   SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
539   attributetype ( BBMRIERICDirectoryCollAttr:8.11
      NAME 'materialStoredTissueFFPE'
541   DESC 'Tissue, formalin fixated and paraffin preserved or equivalent: collection contains
          material of this type (MIABIS-2.0-14).'
      EQUALITY booleanMatch
543   SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
545   attributetype ( BBMRIERICDirectoryCollAttr:8.12
      NAME 'materialStoredImmortalizedCellLines'
547   DESC 'Immortalized cell lines: collection contains material of this type (MIABIS-2.0-14).'
      EQUALITY booleanMatch
549   SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
551   attributetype ( BBMRIERICDirectoryCollAttr:8.13
      NAME 'materialStoredIsolatedPathogen'
553   DESC 'Isolated Pathogen: collection contains material of this type (MIABIS-2.0-14).'
      EQUALITY booleanMatch
555   SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
557
      # Storage temperatures
559
      attributetype ( BBMRIERICDirectoryCollAttr:9.1
561   NAME 'temperatureRoom'
      DESC 'Sample storage temperature - Room temperature - SPREC 2.0 (MIABIS-2.0-15).'
563   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
565
      attributetype ( BBMRIERICDirectoryCollAttr:9.2
567   NAME 'temperature2to10'
      DESC 'Sample storage temperature - between 2 and 10 deg C - SPREC 2.0 (MIABIS-2.0-15).'
569   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
571
      attributetype ( BBMRIERICDirectoryCollAttr:9.3
573   NAME 'temperature-18to-35'
      DESC 'Sample storage temperature - between -18 and -35 deg C - SPREC 2.0 (MIABIS-2.0-15).'
575   EQUALITY booleanMatch
```

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 676550.

32/49

```
577    SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)

       attributetype ( BBMRIERICDirectoryCollAttr:9.4
579     NAME 'temperature-60to-85'
        DESC 'Sample storage temperature - between -60 and -85 deg C - SPREC 2.0 (MIABIS-2.0-15).'
581     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
583
       attributetype ( BBMRIERICDirectoryCollAttr:9.5
585     NAME 'temperatureLN'
        DESC 'Sample storage temperature - liquid nitrogen, -150 to -196 deg C  (MIABIS-2.0-15).'
587     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
589
       attributetype ( BBMRIERICDirectoryCollAttr:9.6
591     NAME 'temperatureOther'
        DESC 'Sample storage temperature - other, text specified (MIABIS-2.0-15).'
593     EQUALITY caseIgnoreMatch
        SUBSTR caseIgnoreSubstringsMatch
595     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

597
       # Collection types
599
       attributetype ( BBMRIERICDirectoryCollAttr:10.1
601     NAME 'collectionTypeCaseControl'
        DESC 'A case-control study design compares two groups of subjects: those with the disease or
            condition under study (cases) and a very similar group of subjects who do not have the
            disease or condition (controls). - EMBL (EFO) - MIABIS-2.0-19.'
603     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
605
       attributetype ( BBMRIERICDirectoryCollAttr:10.2
607     NAME 'collectionTypeCohort'
        DESC 'A form of longitudinal study for the analysis of risk factors following a group of people
            who do not have a disease, and uses correlations to determine the absolute risk of subject
            contraction. - Wikipedia (rewritten) - MIABIS-2.0-19.'
609     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
611
       attributetype ( BBMRIERICDirectoryCollAttr:10.3
613     NAME 'collectionTypeCrossSectional'
        DESC 'A type of observational study that involves data collection from a population, or a
            representative subset, at one specific point in time. - Wikipedia - MIABIS-2.0-19.'
615     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
617
       attributetype ( BBMRIERICDirectoryCollAttr:10.4
619     NAME 'collectionTypeLongitudinal'
        DESC 'Research studies involving repeated observations of the same entity over time. In the
            biobank context, longitudinal studies sample a group of people in a given time period, and
            study them at intervals by the acquisition and analyses of data and/or samples over time. -
            P3G - MIABIS-2.0-19.'
621     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
623
       attributetype ( BBMRIERICDirectoryCollAttr:10.5
625     NAME 'collectionTypeTwinStudy'
        DESC 'Twin studies measure the contribution of genetics (as opposed to environment) to a given
            trait or condition of interest. - MIABIS-2.0-19.'
627     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
629
       attributetype ( BBMRIERICDirectoryCollAttr:10.6
631     NAME 'collectionTypeQualityControl'
        DESC 'A quality control testing study design type is where some aspect of the experiment is
            quality controlled for the purposes of quality assurance. - EMBL (EFO) - MIABIS-2.0-19.'
633     EQUALITY booleanMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
635
       attributetype ( BBMRIERICDirectoryCollAttr:10.7
637     NAME 'collectionTypePopulationBased'
        DESC 'Study done at the population level or among the population groups, generally to find the
            cause, incidence or spread of the disease or to see the response to the treatment,
            nutrition or environment. - Wikipedia (rewritten) - MIABIS-2.0-19.'
```

```
639    EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
641
       attributetype ( BBMRIERICDirectoryCollAttr:10.8
643     NAME 'collectionTypeDiseaseSpecific'
       DESC 'A collection for which material and information is collected from subjects that have
           already developed a particular disease. - EMBL (EFO) - MIABIS-2.0-19.'
645     EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
647
       attributetype ( BBMRIERICDirectoryCollAttr:10.9
649     NAME 'collectionTypeBirthCohort'
       DESC 'A corhort study for which the subjects are followed from the time of birth usually
           including information about gestation and follow up. - MIABIS-2.0-19.'
651     EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
653
       attributetype ( BBMRIERICDirectoryCollAttr:10.10
655     NAME 'collectionTypeOther'
       DESC 'Other type of collection text specified (MIABIS-2.0-19).'
657     EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
659     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

661    # note that clinical collection types are implemented as a separate auxiliary object class in
           order to enforce diagnosisAvailable

663    # This is diagnosis attributes, expected in the ontology:code form (e.g., SNOMED:25*)

665    attributetype ( BBMRIERICDirectoryCollAttr:11
        NAME 'diagnosisAvailable'
667    DESC 'Diagnosis available in the collection, with the ontology prefix, possibly using * and ?
           wildcards, and prefix notation to denote diagnosis nomenclature - so far urn:miriam:icd:
           prefix for ICD-10, and urn:miriam:snomedct: prefix for SNOMED CT (examples being
           urn:miriam:icd:C*, urn:miriam:snomedct:25*) - MIABIS-2.0-17, adapted.'
        EQUALITY caseIgnoreMatch
669     SUBSTR caseIgnoreSubstringsMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
671
       # Collection head
673
       attributetype ( BBMRIERICDirectoryCollAttr:12.1
675     NAME 'collectionHeadFirstName'
       DESC 'First name of a person in charge of the collection.'
677     EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
679     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

681    attributetype ( BBMRIERICDirectoryCollAttr:12.2
        NAME 'collectionHeadLastName'
683    DESC 'Last name of a person in charge of the collection.'
        EQUALITY caseIgnoreMatch
685     SUBSTR caseIgnoreSubstringsMatch
        SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
687
       attributetype ( BBMRIERICDirectoryCollAttr:12.3
689     NAME 'collectionHeadRole'
       DESC 'Official role of the person in charge of the collection: typically PI or Director.'
691     EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
693     SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

695    # Other attributes

697    attributetype ( BBMRIERICDirectoryCollAttr:13.1
        NAME 'collectionSampleAccessFee'
699    DESC 'Denotes whether access to samples may be obtained on fee-based basis.'
        EQUALITY booleanMatch
701     SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)

703    attributetype ( BBMRIERICDirectoryCollAttr:13.2
        NAME 'collectionSampleAccessJointProjects'
705    DESC 'Denotes whether access to samples may be obtained on joint project basis.'
        EQUALITY booleanMatch
707     SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
```

```
709   attributetype ( BBMRIERICDirectoryCollAttr:13.3
       NAME 'collectionSampleAccessDescription'
711    DESC 'Short description of access rules.'
       EQUALITY caseIgnoreMatch
713    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
715
      attributetype ( BBMRIERICDirectoryCollAttr:13.4
717    NAME 'collectionDataAccessFee'
       DESC 'Denotes whether access to data may be obtained on fee-based basis.'
719    EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
721
      attributetype ( BBMRIERICDirectoryCollAttr:13.5
723    NAME 'collectionDataAccessJointProjects'
       DESC 'Denotes whether access to data may be obtained on joint project basis.'
725    EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
727
      attributetype ( BBMRIERICDirectoryCollAttr:13.6
729    NAME 'collectionDataAccessDescription'
       DESC 'Short description of access rules.'
731    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
733    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
735   attributetype ( BBMRIERICDirectoryCollAttr:13.7
       NAME 'collectionSampleAccessURI'
737    DESC 'URI describing access policy for the samples.'
       EQUALITY caseIgnoreMatch
739    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
741
      attributetype ( BBMRIERICDirectoryCollAttr:13.8
743    NAME 'collectionDataAccessURI'
       DESC 'URI describing access policy for the data.'
745    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
747    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')

749   # Size of the collection

751   attributetype ( BBMRIERICDirectoryCollAttr:14.1
       NAME 'collectionOrderOfMagnitude'
753    DESC 'Size of the collection measured as 10^n samples.'
       EQUALITY integerMatch
755    SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

757   attributetype ( BBMRIERICDirectoryCollAttr:14.2
       NAME 'collectionSize'
759    DESC 'Exact size of the collection to the given date.'
       EQUALITY integerMatch
761    SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

763   attributetype ( BBMRIERICDirectoryCollAttr:14.3
       NAME 'collectionSizeTimestamp'
765    DESC 'Date to which the size of the collection was valid, absolute time in ISO 8601 format.'
       EQUALITY integerMatch
767    SYNTAX '1.3.6.1.4.1.1466.115.121.1.27' SINGLE-VALUE)

769
      # Definition of collection object class
771
      objectClass ( BBMRIERICDirectoryCollObj:1
773    NAME 'collection'
       DESC 'Collection (embedded) Object'
775    SUP collaborationStatus
       MUST ( collectionID $ collectionName $ materialStoredDNA $ materialStoredPlasma $
           materialStoredSerum $ materialStoredUrine $ materialStoredSaliva $ materialStoredFaeces $
           materialStoredOther $ materialStoredRNA $ materialStoredBlood $ materialStoredTissueFrozen
           $ materialStoredTissueFFPE $ materialStoredImmortalizedCellLines $
           materialStoredIsolatedPathogen $ collectionTypeCaseControl $ collectionTypeCohort $
           collectionTypeCrossSectional $ collectionTypeLongitudinal $ collectionTypeTwinStudy $
           collectionTypeQualityControl $ collectionTypePopulationBased $
```

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 676550.

35/49

```
           collectionTypeDiseaseSpecific $ collectionTypeBirthCohort $ collectionTypeOther $
           collectionOrderOfMagnitude )
777   MAY ( collectionAcronym $ collectionDescription $  contactIDRef $ contactPriority $
           collectionSexMale $ collectionSexFemale $ collectionSexUnknown $
           collectionSexUndiferrentiated $ collectionAgeLow $ collectionAgeHigh $ collectionAgeUnit $
           collectionAvailableBiologicalSamples $ collectionAvailableSurveyData $
           collectionAvailableImagingData $ collectionAvailableMedicalRecords $
           collectionAvailableNationalRegistries $ collectionAvailableGenealogicalRecords $
           collectionAvailablePhysioBiochemMeasurements $ collectionAvailableOther $ temperatureRoom $
           temperature2to10 $ temperature-18to-35 $ temperature-60to-85 $ temperatureLN $
           temperatureOther $ diagnosisAvailable $ collectionHeadFirstName $ collectionHeadLastName $
           collectionHeadRole $ collectionSampleAccessFee $ collectionSampleAccessJointProjects $
           collectionSampleAccessDescription $ collectionDataAccessFee $
           collectionDataAccessJointProjects $ collectionDataAccessDescription $
           collectionSampleAccessURI $ collectionDataAccessURI $ collectionSize $
           collectionSizeTimestamp $ bioresourceReference $ biobankNetworkIDRef $ geoLatitude $
           geoLongitude ) )


779   objectClass ( BBMRIERICDirectoryCollObj:2
       NAME 'collectionClinical'
781    DESC 'Clinical collection Object'
       AUXILIARY
783    MUST ( diagnosisAvailable ) )


785
      # Biobank network definition
787
      attributetype ( BBMRIERICDirectoryBBNetAttr:1
789    NAME 'biobankNetworkID'
       DESC 'Unique ID of a biobank network within BBMRI-ERIC based on MIABIS 2.0 standard (ISO 3166-1
           alpha-2 + underscore + biobank national ID or name), prefixed with bbmri-eric:bbnet:ID:
           string; if biobank network is on European or higher level, EU_ prefix is to be used instead
           of country prefix.'
791    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
793    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)


795   attributetype ( BBMRIERICDirectoryBBNetAttr:2
       NAME 'biobankNetworkName'
797    DESC 'Biobank network name.'
       EQUALITY caseIgnoreMatch
799    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
801
      attributetype ( BBMRIERICDirectoryBBNetAttr:3
803    NAME 'biobankNetworkAcronym'
       DESC 'Biobank network acronym.'
805    EQUALITY caseIgnoreMatch
       SUBSTR caseIgnoreSubstringsMatch
807    SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)


809   attributetype ( BBMRIERICDirectoryBBNetAttr:4
       NAME 'biobankNetworkDescription'
811    DESC 'Biobank network description.'
       EQUALITY caseIgnoreMatch
813    SUBSTR caseIgnoreSubstringsMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
815
      #
817   # Attributes describing commonalities in biobank networks
      #
819
      attributetype ( BBMRIERICDirectoryBBNetAttr:5.1
821    NAME 'biobankNetworkCommonCollectionFocus'
       DESC 'All the biobanks/collections in the network share the same focus with which the samples
           are collected (e.g., disease specific). Further details of the focus should be provided in
           the biobankNetworkDescription attribute.'
823    EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
825
      attributetype ( BBMRIERICDirectoryBBNetAttr:5.2
827    NAME 'biobankNetworkCommonCharter'
       DESC 'All the biobanks/collections in the network have to have a network charter signed.'
829    EQUALITY booleanMatch
       SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
```

```
831
     attributetype ( BBMRIERICDirectoryBBNetAttr:5.3
833   NAME 'biobankNetworkCommonSOPs'
      DESC 'All the biobanks/collections in the network share the same SOPs.'
835   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
837
     attributetype ( BBMRIERICDirectoryBBNetAttr:5.4
839   NAME 'biobankNetworkCommonDataAccessPolicy'
      DESC 'All the biobanks/collections in the network share the same data access policy.'
841   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
843
     attributetype ( BBMRIERICDirectoryBBNetAttr:5.5
845   NAME 'biobankNetworkCommonSampleAccessPolicy'
      DESC 'All the biobanks/collections in the network share the same sample access policy.'
847   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
849
     attributetype ( BBMRIERICDirectoryBBNetAttr:5.6
851   NAME 'biobankNetworkCommonMTA'
      DESC 'All the biobanks/collections in the network share the same MTA.'
853   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
855
     attributetype ( BBMRIERICDirectoryBBNetAttr:5.7
857   NAME 'biobankNetworkCommonRepresentation'
      DESC 'All the biobanks/collections in the network are represented using the network only.'
859   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
861
     attributetype ( BBMRIERICDirectoryBBNetAttr:5.8
863   NAME 'biobankNetworkCommonURL'
      DESC 'All the biobanks/collections in the network share the same web presentation on the common
          URL.'
865   EQUALITY booleanMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.7' SINGLE-VALUE)
867
     #
869   # Other attributes
     #
871
     attributetype ( BBMRIERICDirectoryBBNetAttr:6
873   NAME 'biobankNetworkURL'
      DESC 'Biobank network URL.'
875   EQUALITY caseIgnoreMatch
      SUBSTR caseIgnoreSubstringsMatch
877   SYNTAX '1.3.6.1.4.1.1466.115.121.1.15' SINGLE-VALUE)
879  attributetype ( BBMRIERICDirectoryBBNetAttr:7
      NAME 'biobankNetworkJuridicalPerson'
881   DESC 'Juristic person of a biobank network according'
      EQUALITY caseIgnoreMatch
883   SUBSTR caseIgnoreSubstringsMatch
      SYNTAX '1.3.6.1.4.1.1466.115.121.1.15')
885
887  # Definition of biobankNetwork object class
889  objectClass ( BBMRIERICDirectoryBBNetObj:1
      NAME 'biobankNetwork'
891   DESC 'BiobankNetwork Object'
      MUST ( biobankNetworkID $ biobankNetworkName $ contactIDRef $ contactPriority $
          biobankNetworkCommonCollectionFocus $ biobankNetworkCommonCharter $
          biobankNetworkCommonSOPs $ biobankNetworkCommonDataAccessPolicy $
          biobankNetworkCommonSampleAccessPolicy $ biobankNetworkCommonMTA $
          biobankNetworkCommonRepresentation $ biobankNetworkCommonURL )
893   MAY ( biobankNetworkAcronym $ biobankNetworkDescription $ biobankNetworkIDRef $ geoLatitude $
          geoLongitude $ biobankNetworkURL $ biobankNetworkJuridicalPerson) )
```
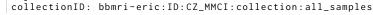
6 Example of Data in LDIF Format

6.1 Example of biobank, collection, and contact information

Standard LDIF file with non-7-bit ASCII values encoded in Base64 encoding:

```
 1  dn: dc=directory,dc=bbmri-eric,dc=eu
    objectClass: domain
 3  objectClass: top
    dc: directory
 5
    dn: ou=contacts,dc=directory,dc=bbmri-eric,dc=eu
 7  objectClass: organizationalUnit
    objectClass: top
 9  ou: contacts
11  dn: ou=biobanknetworks,dc=directory,dc=bbmri-eric,dc=eu
    objectClass: organizationalUnit
13  objectClass: top
    ou: biobanknetworks
15
    dn: ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
17  objectClass: top
    objectClass: organizationalUnit
19  ou: biobanks
21  dn: c=cz,ou=contacts,dc=directory,dc=bbmri-eric,dc=eu
    objectClass: country
23  objectClass: top
    c: cz
25
    dn: c=cz,ou=biobanknetworks,dc=directory,dc=bbmri-eric,dc=eu
27  objectClass: country
    objectClass: top
29  c: cz
31  dn: c=cz,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
    objectClass: country
33  objectClass: top
    c: cz
35
    dn: biobankID=bbmri-eric:ID:CZ_MMCI,c=cz,ou=biobanks,dc=directory,dc=bbmri-eri
37   c,dc=eu
    biobankID: bbmri-eric:ID:CZ_MMCI
39  biobankName: Bank of Biological Material, Masaryk Memorial Cancer Institute
    biobankJuridicalPerson: Masaryk Memorial Cancer Institute
41  biobankCountry: CZ
    objectClass: biobank
43  objectClass: biobankClinical
    biobankPartnerCharterSigned: FALSE
45  biobankAcronym: MMCI
    biobankURL: http://www.recamo.cz/en/bbmri/
47  biobankDescription:: QmlvYmFuayBhdCBNTUNJIHdhcyBlc3RhYmxpc2hlZCB0byBwcmVzZXJ2Z
     SBwYXRpZW50LWRlcml2ZWQgc3BlY2ltZW5zIHN1Y2ggYXMgdHVtb3VyyBhbmQgYmxvb2QtZGVyaX
49  ZlZCBzYW1wbGVzIC4gSXRzIHN0cnVjdHVyZSBjb21wcmlzZXMgb2YgYSAibG9uZy10ZXJtIHN0b3J
     hZ2UgcmVwb3NpdG9yeSIgZGVzaWduZWQgdG8gc3RvcmUgYSB2aXJpZW5zIHNldCBvZiBw
51  YXRpZW50IHNhbXBsZXMgc3VjaCBhcyB0dW1vdXJzLCBhZGphY2VudCB0aXNzdWVzLCBnZW5vbWljI
     EROQSBhbmQgc2VydW0gYXQgc3VyZ2VyeSwgYW5kIHRoZSAic2hvcnQgdGVybSBzdG9yYWdlIHJlcG
53  9zaXRvcnkiIGRlc2lnbmVkIHRvIHN0b3JlIHNlcnVtIGFscXVvdHMgZnJvb20gcGVyaW9kaWMgZGV
     0ZXJtaW5hdGlvbiBvZiBzb21x1YmxlIGJsb29kIHR1bW91ciBtYXJrZXJzLiAgQ29tcHJlaGVuc2l2
55  ZSBjbGluaWNhbCBkYXRhIGFyZSBhdmFpbGFibGUgdGhyb3VnaCB0aGUgaG9zcGl0YWwgaW5mcm9tY
     XRpb24gc3lzdGVtIHRvIGVuYWJsZSBmdXJ0aGVyIHRyYW5zbGF0aW9uYWwgYW5kIGNsaW5pY2FsIH
57  Jlc2VhcmNoLiA=
    geoLatitude: 49.19426448
59  geoLongitude: 16.5890801
    contactIDRef: bbmri-eric:contact:CZ_MMCI
61  contactPriority: 2
63  dn: collectionID=bbmri-eric:ID:CZ_MMCI:collection:all_samples,biobankID=bbmri-
     eric:ID:CZ_MMCI,c=cz,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
65  objectClass: collection
    collectionID: bbmri-eric:ID:CZ_MMCI:collection:all_samples
```

```
67   collectionName: Main collection of Bank of Biological Material, Masaryk Memori
     al Cancer Institute, comprising of all the samples.
69   contactIDRef: bbmri-eric:contact:CZ_MMCI
     contactPriority: 1
71   materialStoredDNA: TRUE
     materialStoredPlasma: FALSE
73   materialStoredSerum: TRUE
     materialStoredUrine: FALSE
75   materialStoredSaliva: FALSE
     materialStoredFaeces: FALSE
77   materialStoredOther: FALSE
     materialStoredRNA: FALSE
79   materialStoredBlood: TRUE
     materialStoredTissueFrozen: TRUE
81   materialStoredTissueFFPE: TRUE
     materialStoredImmortalizedCellLines: FALSE
83   materialStoredIsolatedPathogen: FALSE
     collectionAvailableBiologicalSamples: TRUE
85   collectionAvailableSurveyData: FALSE
     collectionAvailableImagingData: FALSE
87   collectionAvailableMedicalRecords: TRUE
     collectionAvailableNationalRegistries: FALSE
89   collectionAvailableGenealogicalRecords: FALSE
     collectionAvailablePhysioBiochemMeasurements: TRUE
91   collectionAvailableOther: FALSE
     collectionSampleAccessFee: FALSE
93   collectionSampleAccessJointProjects: TRUE
     collectionSampleAccessDescription: Further access details available upon reque
95    st.
     collectionDataAccessFee: FALSE
97   collectionDataAccessJointProjects: TRUE
     collectionDataAccessDescription: Further access details available upon request
99    .
     collectionSampleAccessURI: http://www.recamo.cz/en/bbmri/
101  collectionDataAccessURI: http://www.recamo.cz/en/bbmri/
     collectionSexMale: TRUE
103  collectionSexFemale: TRUE
     collectionTypeCaseControl: FALSE
105  collectionTypeCohort: FALSE
     collectionTypeCrossSectional: FALSE
107  collectionTypeLongitudinal: FALSE
     collectionTypeTwinStudy: FALSE
109  collectionTypeQualityControl: FALSE
     collectionTypePopulationBased: FALSE
111  collectionTypeDiseaseSpecific: FALSE
     collectionTypeBirthCohort: FALSE
113  collectionTypeOther: FALSE
     diagnosisAvailable: urn:miriam:icd:D*
115  diagnosisAvailable: urn:miriam:icd:C*
     collectionOrderOfMagnitude: 4
117
     dn: contactID=bbmri-eric:contact:CZ_MMCI,c=cz,ou=contacts,dc=directory,dc=bbmr
119   i-eric,dc=eu
     contactID: bbmri-eric:contact:CZ_MMCI
121  objectClass: contactInformation
     contactEmail: nenutil@mou.cz
123  contactCountry: CZ
     contactFirstName: Rudolf
125  contactLastName: Nenutil
     contactPhone: +420543133411
127  contactAddress:: xb1sdXTDvSBrb3BlYyA3
     contactZIP: 653 53
129  contactCity: Brno
```

The same data decoded^p from Base64 for readability reasons:

---

^{p} The conversion can be implemented, e.g., as follows:

```
cat input.ldif | \
   perl -MMIME::Base64 -MEncode=decode -n -00 \
   -e 's/\n//g;s/(?<=:: )(\S+)/decode("UTF-8",decode_base64($1))/eg;print'
```

```
 1  dn: dc=directory,dc=bbmri-eric,dc=eu
    objectClass: domain
 3  objectClass: top
    dc: directory
 5
    dn: ou=contacts,dc=directory,dc=bbmri-eric,dc=eu
 7  objectClass: organizationalUnit
    objectClass: top
 9  ou: contacts
11  dn: ou=biobanknetworks,dc=directory,dc=bbmri-eric,dc=eu
    objectClass: organizationalUnit
13  objectClass: top
    ou: biobanknetworks
15
    dn: ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
17  objectClass: top
    objectClass: organizationalUnit
19  ou: biobanks
21  dn: c=cz,ou=contacts,dc=directory,dc=bbmri-eric,dc=eu
    objectClass: country
23  objectClass: top
    c: cz
25
    dn: c=cz,ou=biobanknetworks,dc=directory,dc=bbmri-eric,dc=eu
27  objectClass: country
    objectClass: top
29  c: cz
31  dn: c=cz,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
    objectClass: country
33  objectClass: top
    c: cz
35
    dn: biobankID=bbmri-eric:ID:CZ_MMCI,c=cz,ou=biobanks,dc=directory,
37   dc=bbmri-eric,dc=eu
    biobankID: bbmri-eric:ID:CZ_MMCI
39  biobankName: Bank of Biological Material, Masaryk Memorial Cancer Institute
    biobankJuridicalPerson: Masaryk Memorial Cancer Institute
41  biobankCountry: CZ
    objectClass: biobank
43  objectClass: biobankClinical
    biobankPartnerCharterSigned: FALSE
45  biobankAcronym: MMCI
    biobankURL: http://www.recamo.cz/en/bbmri/
47  biobankDescription:: Biobank at MMCI was established to preserve patient-derived specimens such
        as tumours and blood-derived samples . Its structure comprises of a "long-term storage
        repository" designed to store a comprehensive set of patient samples such as tumours,
        adjacent tissues, genomic DNA and serum at surgery, and the "short term storage repository"
        designed to store serum aliquots from periodic determination of soluble blood tumour
        markers.  Comprehensive clinical data are available through the hospital infromation system
        to enable further translational and clinical research.
    geoLatitude: 49.19426448
49  geoLongitude: 16.5890801
    contactIDRef: bbmri-eric:contact:CZ_MMCI
51  contactPriority: 2
53  dn: collectionID=bbmri-eric:ID:CZ_MMCI:collection:all_samples,
     biobankID=bbmri-eric:ID:CZ_MMCI,c=cz,ou=biobanks,dc=directory,
55   dc=bbmri-eric,dc=eu
    objectClass: collection
57  collectionID: bbmri-eric:ID:CZ_MMCI:collection:all_samples
    collectionName: Main collection of Bank of Biological Material, Masaryk Memorial Cancer
        Institute, comprising of all the samples.
59  contactIDRef: bbmri-eric:contact:CZ_MMCI
    contactPriority: 1
61  materialStoredDNA: TRUE
    materialStoredPlasma: FALSE
63  materialStoredSerum: TRUE
    materialStoredUrine: FALSE
65  materialStoredSaliva: FALSE
    materialStoredFaeces: FALSE
67  materialStoredOther: FALSE
```

```
69   materialStoredRNA: FALSE
     materialStoredBlood: TRUE
     materialStoredTissueFrozen: TRUE
71   materialStoredTissueFFPE: TRUE
     materialStoredImmortalizedCellLines: FALSE
73   materialStoredIsolatedPathogen: FALSE
     collectionAvailableBiologicalSamples: TRUE
75   collectionAvailableSurveyData: FALSE
     collectionAvailableImagingData: FALSE
77   collectionAvailableMedicalRecords: TRUE
     collectionAvailableNationalRegistries: FALSE
79   collectionAvailableGenealogicalRecords: FALSE
     collectionAvailablePhysioBiochemMeasurements: TRUE
81   collectionAvailableOther: FALSE
     collectionSampleAccessFee: FALSE
83   collectionSampleAccessJointProjects: TRUE
     collectionSampleAccessDescription: Further access details available upon request.
85   collectionDataAccessFee: FALSE
     collectionDataAccessJointProjects: TRUE
87   collectionDataAccessDescription: Further access details available upon request.
     collectionSampleAccessURI: http://www.recamo.cz/en/bbmri/
89   collectionDataAccessURI: http://www.recamo.cz/en/bbmri/
     collectionSexMale: TRUE
91   collectionSexFemale: TRUE
     collectionTypeCaseControl: FALSE
93   collectionTypeCohort: FALSE
     collectionTypeCrossSectional: FALSE
95   collectionTypeLongitudinal: FALSE
     collectionTypeTwinStudy: FALSE
97   collectionTypeQualityControl: FALSE
     collectionTypePopulationBased: FALSE
99   collectionTypeDiseaseSpecific: FALSE
     collectionTypeBirthCohort: FALSE
101  collectionTypeOther: FALSE
     diagnosisAvailable: urn:miriam:icd:D*
103  diagnosisAvailable: urn:miriam:icd:C*
     collectionOrderOfMagnitude: 4
105
     dn: contactID=bbmri-eric:contact:CZ_MMCI,c=cz,ou=contacts,
107   dc=directory,dc=bbmri-eric,dc=eu
     contactID: bbmri-eric:contact:CZ_MMCI
109  objectClass: contactInformation
     contactEmail: nenutil@mou.cz
111  contactCountry: CZ
     contactFirstName: Rudolf
113  contactLastName: Nenutil
     contactPhone: +420543133411
115  contactAddress:: Žlutý kopec 7
     contactZIP: 653 53
117  contactCity: Brno
```

6.2 Example of referral

```
     dn: c=ee,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
2    objectClass: referral
     objectClass: extensibleObject
4    c: ee
     ref: ldap://193.40.12.248/c=ee,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu
6
     dn: c=ee,ou=biobanknetworks,dc=directory,dc=bbmri-eric,dc=eu
8    objectClass: referral
     objectClass: extensibleObject
10   c: ee
     ref: ldap://193.40.12.248/c=ee,ou=biobanknetworks,dc=directory,dc=bbmri-eric,dc=eu
12
     dn: c=ee,ou=contacts,dc=directory,dc=bbmri-eric,dc=eu
14   objectClass: referral
     objectClass: extensibleObject
16   c: ee
     ref: ldap://193.40.12.248/c=ee,ou=contacts,dc=directory,dc=bbmri-eric,dc=eu
```

7 Recommended Configuration of OpenLDAP Server

7.1 OpenLDAP Cookbook for National Node

This is a very simple deployment guide which should be valid for Ubuntu/Debian distros. It is intended to be simplest config for the BBMRI-ERIC national directory, without any fancy stuff nor integration with any existing systems.

1. install OpenLDAP server and take it down for configuration

   ```
   apt-get install slapd
   /etc/init.d/slapd stop
   ```

2. get your TLS/SSL key and certs ready – in the worst case if you don't have access to commonly accepted certificates (such as ones provided by Terena Certificate Service), just generate a snakeoil ones:

   ```
   mkdir /etc/ldap/ssl
   cd /etc/ldap/ssl
   openssl genrsa -out cert.key 1024
   openssl req -new -key cert.key -out cert.csr
   openssl x509 -req -in cert.csr -days 4096 -signkey cert.key -out cert.crt
   chown openldap:openldap *
   chmod 600 cert.key
   ```

3. enable LDAPS in /etc/default/slapd

   ```
   SLAPD_SERVICES="ldap:/// ldapi:/// ldaps:///
   ```

4. remove the default config directory out of the way

   ```
   mv /etc/ldap/slapd.d /etc/ldap/slapd.d-default
   ```

5. create new config file (we will use this for simplicity reasons, it can be easily converted to the more modern LDAP-based config later on) as shown in Section 7.2.2.

   What this config does is

   a) sets up a OpenLDAP server with all the schemas we need at the moment (and even a few we may need in the future)

   b) enables server to listen at ports 389 (plain LDAP) and 636 (LDAP with SSL/TLS)

   c) requires authentication to use SSL/TLS

   d) creates local database, in this case for the bbmri.cz

   e) requires authentication in order to enable writes into the database, while reads are also supported for anonymous users over unencrypted connection

6. put the biobank.schema into `/etc/ldap/schema/`

7. start OpenLDAP slapd server

   ```
   /etc/init.d/slapd start
   ```

8. if something goes wrong, you can debug the slapd by running it with `-d -1` in the command line:

   ```
   /usr/sbin/slapd -g openldap -u openldap -f /etc/ldap/slapd.conf -d -1 -h "ldap:///
       ldapi:/// ldaps:///"
   ```

9. you can fill the data in using slapadd (you need to stop slapd first in this case) or ldapadd

10. inform BBMRI-ERIC headquarters (or CS IT) where your server is running and which domain it is serving; we will add appropriate referral into the `directory.bbmri-eric.eu` so that your biobanks can be looked up from it

11. once everything is done, try searching through the directory:

```
ldapsearch -v -x -h directory.bbmri-eric.eu -p 389 -b 'dc=directory,dc=bbmri-eric,dc=eu' -s
    sub -C
```

(-C is important at the moment to make ldapsearch from OpenLDAP 2.x chase (= follow) referrals if in place) or using LDAP Admin tool[q]

12. you can also test your own server in a similar way

7.2 `slapd.conf`

For simplicity reasons, we use an older approach to configuring OpenLDAP server using as single `slapd.conf` configuration file.

7.2.1 Central BBMRI-ERIC server with chaining support

```
 1  # enterprise directory
    include         /etc/ldap-devel/schema/core.schema
 3  include         /etc/ldap-devel/schema/cosine.schema
    include         /etc/ldap-devel/schema/inetorgperson.schema
 5  include         /etc/ldap-devel/schema/nis.schema
    include         /etc/ldap-devel/schema/openldap.schema
 7
    # biobank specifc schema
 9  include         /etc/ldap-devel/schema/biobank.schema
    include         /etc/ldap-devel/schema/eduPerson201310.schema
11
    # modules
13  modulepath      /usr/lib/ldap/
    moduleload      back_bdb
15  moduleload      back_hdb
    moduleload      back_ldap
17
    # runtime options
19  pidfile         /var/run/slapd-devel/slapd.pid
    argsfile        /var/run/slapd-devel/slapd.args
21
    # TLS
23  TLSCertificateFile /etc/ldap-devel/ssl-new/servercert.pem
    TLSCertificateKeyFile /etc/ldap-devel/ssl-new/serverkey.pem
25  TLSCACertificateFile /etc/ldap-devel/ssl-new/chain_TERENA_SSL_CA_3.pem
27  sizelimit 10000
    timelimit 60
29
    access to
       attrs=biobankITSupportAvailable,biobankITStaffSize,biobankISAvailable,biobankHISAvailable
31        by users write
          by * none
33
    access to *
35        by users write
          by * read
37        by peername.regex=10\..*\..*\..* write
```

---

[q] `http://www.ldapadmin.org/`

```
39           by peername.regex=127\..*\..*\..* write

     # biobank database definition
41   database         hdb
     suffix           "dc=directory,dc=bbmri-eric,dc=eu"
43   rootdn           "cn=root,dc=directory,dc=bbmri-eric,dc=eu"
     rootpw           {SSHA}**PASSWORDREMOVED**
45   directory        /var/lib/ldap-devel
     #index            objectClass    pres,eq
47   index            default pres,eq,sub,approx

49

     overlay          chain
51   chain-max-depth 2
     chain-return-error       FALSE
```

7.2.2 Example of National Node server

```
     # enterprise directory
2    include          /etc/ldap/schema/core.schema
     include          /etc/ldap/schema/cosine.schema
4    include          /etc/ldap/schema/inetorgperson.schema
     include          /etc/ldap/schema/nis.schema
6    include          /etc/ldap/schema/openldap.schema

8    # biobank specifc schema
     include          /etc/ldap/schema/biobank.schema
10
     # modules
12   modulepath       /usr/lib/ldap/
     moduleload       back_bdb
14   moduleload       back_hdb

16   # runtime options
     pidfile          /var/run/slapd/slapd.args
18   argsfile         /var/run/slapd/slapd.pid

20   # TLS
     security simple_bind=128
22   TLSCertificateFile /etc/ldap/ssl/www.bbmri.cz-1376563249.cer
     TLSCertificateKeyFile /etc/ldap/ssl/bbmri.cz.key.pem
24   TLSCACertificateFile /etc/ldap/ssl/chain_TERENA_SSL_CA_2.pem

26
     # biobank database definition
28   database         hdb
     suffix           "c=cz,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu"
30   rootdn           "CN=root,c=cz,ou=biobanks,dc=directory,dc=bbmri-eric,dc=eu"
     rootpw           {SSHA}**PASSWORDREMOVED**
32   directory        /var/lib/ldap
     index            objectClass    pres,eq
34
     access to
         attrs=biobankITSupportAvailable,biobankITStaffSize,biobankISAvailable,biobankHISAvailable
36       by users write
         by * none
38
     access to *
40       by ssf=128 users write
         by users write
42       by * read
     #       by peername.regex=10\..*\..*\..* write
```

7.3 Configuration of backend databases

```
   set_cachesize 0 2097152 0
2  set_lk_max_objects 1500
   set_lk_max_locks 1500
4  set_lk_max_lockers 1500
   set_lg_regionmax        1048576
6  set_lg_max              10485760
   set_lg_bsize            2097152
```

8 Changes from Directory 1.0 to 2.0

This section is intended for ADOPT BBMRI-ERIC project to clarify development of the Directory toward Deliverable D3.1.

Directory 1.0 (released in July 2015) has been developed as a pre-ADOPT tool, while Directory 2.0 (released in December 2015) has been implemented within the ADOPT project.

8.1 Change of Data Model

Directory 1.0 used a simplified model based on pre-release MIABIS 2.0: because of uncertainties in semantics of collection entity at the time of release of Directory 1.0, the biobank and collection were merged into a single biobank entity.

For Directory 2.0, the following changes were implemented:

- Introduction of collection entity:

    - samples and data related information was moved from biobank entity to the collection entity,

    - collections were defined as partitions of a parent set (a biobank or a collection), where each countable element (typically sample, but can be also extended to images for imaging biobanks, etc.) participates in exactly one collection on the given level of recursion,

    - partitioning-based collections follow recommendation of the MIABIS 2.0,[27]

    - biobank entity now covers only attributes related to the institutional aspects of biobanks.

- Introduction of standalone auxiliary contact information entity, which is a result of normalization of the data model after introduction of collection entity, as several collections may share the same contact information. This concept has been propagated also to biobank and biobank network entities.

- Addition of head person of biobanks, collection, or biobank network: this extends previously available contact information.

- Introduction of biobank networks with *m:n* mapping between biobanks/collections and biobank networks.

- Support for improved estimates of size of biobanks: beyond the mandatory order of magnitude, the collections can be reported with exact number of samples stored and a time stamp (note the uncertainty related to sample/aliquot definition discussed earlier in the paper).

8.2 Change of User Interfaces

Directory 2.0 user interface based on LifeRay content managements system (used for managing BBMRI-ERIC website) has been updated to reflect changes in the data model described in the previous section.

The update of the user interface has also been used to improve search in diagnosis field, which now supports limited semantic search for ICD-10 codes (searching for more generic as well as more specialized diagnoses relative to the user-requested one(s)).

8.3 Data Curation

Considerable effort has been invested into updating data to the new structure of Directory 2.0. In the first stage, this upgrade has been automated by creating one collection of all the samples in each biobanks, in order to facilitate the upgrade process. Afterwards the National Nodes have been invited to update their data, in order to improve their value for the users: introducing their existing collections and biobank networks. Another important aspect was review of existing standalone collections, if the new semantics of biobanks and collections allows for embedding standalone collections into biobanks (previously the standalone collections were used for two purposes: to describe collections that stand outside of institutionalized biobanks in the given country, and collections that required their own visibility in the Directory 1.0 – the second reason is no longer valid in Directory 2.0 as collections are already visible for the Directory users).

Data curation effort has also led to reimplementation of the data quality check tool, bringing it in line with updated data structures, as well as implementing new quality checks specific to the updated structure (e.g. checks for broken links between entities, checks of newly introduced attributes).

References

1. Harris, J. R., Burton, P., Knoppers, B. M., Lindpaintner, K., Bledsoe, M., Brookes, A. J., Budin-Ljøsne, I., Chisholm, R., Cox, D., Deschênes, M., *et al.* Toward a roadmap in global biobanking for health. *European Journal of Human Genetics* **20,** 1105–1111 (2012).

2. Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., van Dijk, F., van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H. & Stolk, R. P. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* **44,** 1172–1180 (Aug. 2015).

3. Litton, J.-E. in *Methods in Biobanking* 343–361 (Springer, 2011).

4. Wichmann, H.-E., Kuhn, K. A., Waldenberger, M., Schmelcher, D., Schuffenhauer, S., Meitinger, T., Wurst, S. H., Lamla, G., Fortier, I., Burton, P. R., *et al.* Comprehensive catalog of European biobanks. *Nature Biotechnology* **29,** 795–797 (2011).

5. Yuille, M., van Ommen, G.-J., Bréchot, C., Cambon-Thomsen, A., Dagher, G., Landegren, U., Litton, J.-E., Pasterk, M., Peltonen, L., Taussig, M., *et al.* Biobanking for europe. *Briefings in bioinformatics* **9,** 14–24 (2008).

6. Ioannidis, J. P., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., Falchi, M., Furlanello, C., Game, L., Jurman, G., *et al.* Repeatability of published microarray gene expression analyses. *Nature genetics* **41,** 149–155 (2009).

7. Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature reviews Drug discovery* **10,** 712 (2011).

8. Begley, C. G. & Ellis, L. M. Drug development: Raise standards for preclinical cancer research. *Nature* **483,** 531–533 (2012).

9. Bissell, M. Reproducibility: The risks of the replication drive. *Nature* **503,** 333–334 (2013).

10. Hansson, M. G., Dillner, J., Bartram, C. R., Carlson, J. A. & Helgesson, G. Should donors be allowed to give broad consent to future biobank research? *The Lancet Oncology* **7,** 266–269 (2006).

11. Bäumen, T. S. I. D., Paci, D. & Ibarreta, D. Data Protection and Sample Management in Biobanking–A legal dichotomy. *Genomics, Society & Policy* **6** (2010).

12. Williams, H., Spencer, K., Sanders, C., Lund, D., Whitley, E. A., Kaye, J. & Dixon, W. G. Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR medical informatics* **3** (2015).

13. Chadwick, R. & Strange, H. Harmonisation and standardisation in ethics and governance: conceptual and practical challenges. *The Governance of Genetic Information: Who Decides* **9,** 201–213 (2009).

14. European medical research escapes stifling privacy laws. *Nature.* doi:`10.1038/nature.2015.19054` (Dec. 2015).

15. Verissimo, P. E. & Bessani, A. E-biobanking: What Have You Done to My Cell Samples? *IEEE Security & Privacy,* 62–65 (2013).

16. Gholami, A., Lind, A.-S., Reichel, J., Litton, J.-E., Edlund, A. & Laure, E. Privacy Threat Modeling for Emerging BiobankClouds. *Procedia Computer Science* **37,** 489–496 (2014).

17. Gholami, A., Dowling, J. & Laure, E. *A security framework for population-scale genomics analysis* in *High Performance Computing & Simulation (HPCS), 2015 International Conference on* (2015), 106–114.

18. Knoppers, B., Fortier, I., Legault, D. & Burton, P. The Public Population Project in Genomics (P3G): a proof of concept. *Eur J Hum Genet* **16,** 664–665 (2008).

19. Thompson, R., Johnston, L., Taruscio, D., Monaco, L., Béroud, C., Gut, I. G., Hansson, M. G., 't Hoen, P.-B. A., Patrinos, G. P., Dawkins, H., Ensini, M., Zatloukal, K., Koubi, D., Heslop, E., Paschall, J. E., Posada, M., Robinson, P. N., Bushby, K. & Lochmüller, H. RD-Connect: An Integrated Platform Connecting Databases, Registries, Biobanks and Clinical Bioinformatics for Rare Disease Research. *Journal of General Internal Medicine* **29,** 780–787. ISSN: 1525-1497 (2014).

20. Foster, I., Kesselman, C. & Tuecke, S. The anatomy of the grid: Enabling scalable virtual organizations. *International journal of high performance computing applications* **15,** 200–222 (2001).

21. Czajkowski, K., Fitzgerald, S., Foster, I. & Kesselman, C. *Grid information services for distributed resource sharing* in *High Performance Distributed Computing, 2001. Proceedings. 10th IEEE International Symposium on* (2001), 181–194.

22. Plale, B., Dinda, P. & von Laszewski, G. *Key concepts and services of a grid information service* in *Proceedings of the 15th International Conference on Parallel and Distributed Computing Systems (PDCS 2002)* (2002), 437–442.

23. Andrzejak, A. & Xu, Z. *Scalable, efficient range queries for grid information services* in *Peer-to-Peer Computing, 2002.(P2P 2002). Proceedings. Second International Conference on* (2002), 33–40.

24. Cai, M., Frank, M., Chen, J. & Szekely, P. Maan: A multi-attribute addressable network for grid information services. *Journal of Grid Computing* **2,** 3–14 (2004).

25. Puppin, D., Moncelli, S., Baraglia, R., Tonellotto, N. & Silvestri, F. in *Euro-Par 2005 Parallel Processing* 454–464 (Springer, 2005).

26. Puchois, P. *General overview of industry needs and requirements* in *HandsOn: Biobanks* (July 2015).

27. Merino-Martinez, R., Norlin, L., van Enckevort, D., Anton, G., Schuffenhauer, S., Silander, K., Mook, L., Holub, P., Bild, R., Swertz, M. & Litton, J.-E. Towards Global Biobank Integration by Implementation of the Minimum Information About Biobank data Sharing (MIABIS 2.0 Core). *Biopreservation and Biobanking.* ISSN: 1947-5535. doi:`10.1089/bio.2015.0070` (Mar. 2016).

28. Norlin, L., Fransson, M. N., Eriksson, M., Merino-Martinez, R., Anderberg, M., Kurtovic, S. & Litton, J.-E. A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreservation and biobanking* **10,** 343–348 (2012).

29. Zeilenga, K. *Lightweight directory access protocol (LDAP): Technical specification road map* RFC 4510 (IETF, June 2006), 1–7. <https://www.rfc-editor.org/rfc/rfc4510.txt>.

30. Swertz, M. A., Dijkstra, M., Adamusiak, T., van der Velde, J. K., Kanterakis, A., Roos, E. T., Lops, J., Thorisson, G. A., Arends, D., Byelas, G., Muilu, J., Brookes, A. J., de Brock, E. O., Jansen, R. C. & Parkinson, H. The MOLGENIS toolkit: rapid prototyping of biosoftware at the push of a button. *BMC bioinformatics* **11,** 1–9. ISSN: 1471-2105 (2010).

31. Pautasso, C., Wilde, E. & Alarcon, R. *REST: Advanced Research Topics and Practical Applications* (Springer, 2014).