



BBMRI-ERIC

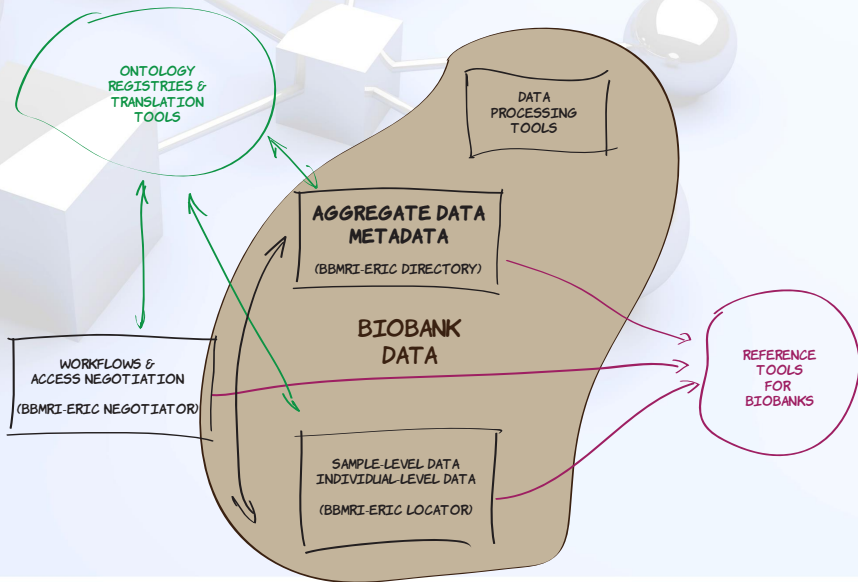
Biobanking and
BioMolecular resources
Research Infrastructure

Metadata Services of BBMRI-ERIC: Directory 1.0 and 2.0

Assoc. Prof. RNDr. Petr Holub, Ph.D.
IT & Data Protection Manager @ BBMRI-ERIC
CIO of BBMRI-ERIC CS IT

4th National Biobank Symposium
Berlin, 2015–12–09

Metadata Services of BBMRI-ERIC



BBMRI-ERIC Directory Use Cases

- ▶ A **researcher wants to identify biobanks** of his/her interest: samples, data, access conditions, etc.
 - retrieving samples/data,
 - hosting samples/data.

BBMRI-ERIC Directory Use Cases

- ▶ Limitation of aggregate data: **What biobanks might have samples/data of my interest?**
 - inability to ask AND questions
 - ⇒ no guarantee there is an actual combination of parameters,
 - samples/data may not be usable for the given analytical methods,
 - samples/data may come with an informed consent that does not allow their use for a particular research purpose.

⇒ **Directory allows for filtering out** biobanks that certainly do not have material of interest.

- input for other tools (e.g., BBMRI-ERIC Sample/Data Negotiator),
- can be used for manual interactions with biobanks.

BBMRI-ERIC Directory Use Cases

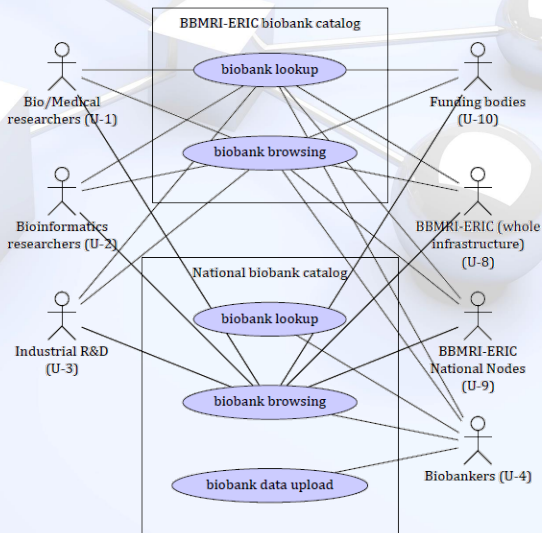
- ▶ A **biobanker** wants to
 - **promote visibility** of their biobank,
 - including all their services: samples + data + expertise,
 - promote access to samples and data,
 - **search for similar** biobanks,
 - **minimize overhead** resulting from integration of their biobank with BBMRI-ERIC
 - e.g., manual data input,
 - most **biobanks lack** an on-line **interface** (= connector) to their system **to publish such information**.

- ▶ BBMRI-ERIC **National Nodes** want to **promote visibility** of their biobanks and services.

BBMRI-ERIC Directory Use Cases

- ▶ **Participants (patients/donors)** want to
 - know **where their samples/data** are hosted,
 - for what research (or what good in general) **their samples/data** are used.
- ▶ **Funding agencies** want to **present the extent of infrastructures** they are funding.

BBMRI-ERIC Directory Use Cases



Directory 1.0

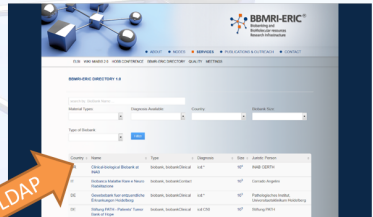
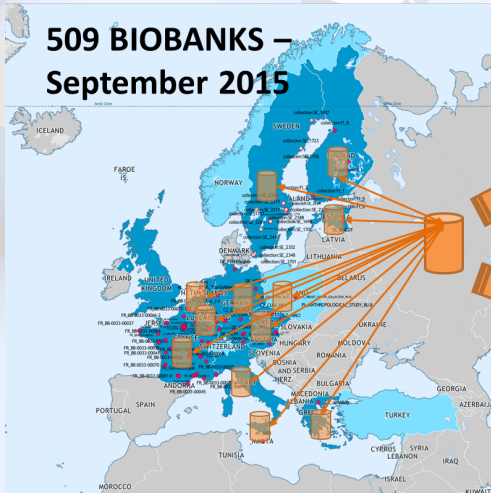
<http://bbmri-eric.eu/bbmri-eric-directory>

- ▶ Released: July 2015
- ▶ Based on flattened **MIABIS 2.0**¹
 - merged biobanks and collections for simplicity.
- ▶ **Machine-readable and human readable interfaces**
 - machine readable: LDAP, JSON,
 - human readable: web-based.
- ▶ **Federated architecture**
 - hierarchy of LDAP servers,
 - optional other interfaces for National Nodes that do not want to run their own LDAP (e.g. JSON, CSV),
 - storing data in BBMRI-ERIC LDAP server for National Nodes that do not want to run any infrastructure.

¹<https://github.com/MIABIS/miabis/wiki>

Directory 1.0

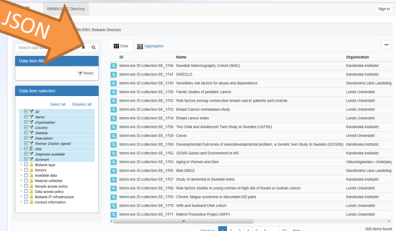
On BBMRI-ERIC web:
<http://www.bbmri-eric.eu/web/guest/bbmri-eric-directory-1.0>



LDAP

JSON

2 USER INTERFACES



Directory 1.0

► **Mandatory information:**

- institutional aspects
 - ID, juridical person, country
 - signed BBMRI-ERIC charter
- basic contact information
 - email
- material types
 - DNA, plasma, serum, urine, saliva, feces, RNA, blood, tissue frozen, FFPE or other fixed tissue, immortalized cell lines, isolated pathogen, other
- available diagnoses
 - preferably in ICD-10, support for other ontologies
 - mandatory only for clinical biobanks

Directory 1.0

▶ **Optional information:**

- description & URL of biobank,
- detailed contact information
 - person, phone, address, geo-coordinates,
- access conditions for samples and data separately
 - joint-project based, fee-based, other,
 - URL describing policies,
- available data types
 - male samples data, female samples data, biological samples, survey data, imaging data, medical records, national registries, genealogical records, physio biochem measurements, other.

Types of Biobanks in Directory

► Clinical biobanks

- 136 in BBMRI-ERIC now, estimated 11.6 M samples,
- problem with very heterogeneous and even unstructured data.



Source: BBMRI-ERIC Directory 1.0

Types of Biobanks in Directory

► Population biobanks

- 189 in BBMRI-ERIC now (incl. research studies), estimated 15.4 M samples,
- tend to have more homogeneous/consistent datasets.



Source: BBMRI-ERIC Directory 1.0

Types of Biobanks in Directory

- ▶ **Other/specific types of biobanks**
 - 2 veterinary biobanks
 - ... can still serve human medicine research
 - biobanks containing fetal samples
- ▶ Is biobank without biological samples (= “dry biobank”) also a biobank?
 - population datasets (e.g., national registries),
 - imaging (e.g., radiology, pathology),
 - definitely of interest, maybe just renaming them to “databanks”.

Preview of Directory 2.0

- ▶ Directory 2.0 introduces new entities:
 - collections,
 - biobanks now only represent the institutional aspects,
 - collections contain all the sample-related information,
 - biobank networks,
 - contact information,
 - to avoid contact duplication for biobanks, collections, and biobank networks.

Preview of Directory 2.0

dc=directory,dc=bbmri-eric,dc=eu [webinar.bbmri-eric.eu]

- + ou=biobanknetworks
- + ou=biobanks
 - + c=at
 - + c=be
 - + c=cz
 - + biobankID=bbmri-eric:ID:CZ_CUNI_HK
 - + biobankID=bbmri-eric:ID:CZ_CUNI_LF1
 - + biobankID=bbmri-eric:ID:CZ_CUNI_PILS
 - biobankID=bbmri-eric:ID:CZ_MMCI
 - collectionID=bbmri-eric:ID:CZ_MMCI:collection:all_samples
 - + collectionID=bbmri-eric:ID:CZ_MMCI:collection:all_samples:LTS
 - + collectionID=bbmri-eric:ID:CZ_MMCI:collection:all_samples:STS
 - + biobankID=bbmri-eric:ID:CZ_UPOL_LF

⋮

- + c=se
- ou=contacts
 - + c=at
 - + c=be
 - + c=cz
 - contactID=bbmri-eric:contact:CZ_CUNI_HK
 - contactID=bbmri-eric:contact:CZ_CUNI_LF1
 - contactID=bbmri-eric:contact:CZ_CUNI_PILS

BIOBANKS (indicated by a red arrow pointing to the CZ_MMCI biobank entry)
COLLECTIONS (indicated by a red arrow pointing to the all_samples collection entry)

Preview of Directory 2.0

ID: bbmri-eric:ID:CZ_MMCI



Bank of Biological Material, Masaryk Memorial Cancer Institute

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

Head:
Institution:

BIOBANK



☛ Main collection of Bank of Biological Material, Masaryk Memorial Cancer Institute, comprising of all the samples.

☐ Long Term Storage Collection

☐ Short Term Storage Collection

COLLECTIONS

LONG TERM STORAGE COLLECTION

MAIN CONTACT

Name: Rudolf Nenutil
 E-Mail: nenutil@mou.cz
 Address: Zluty kopec 7
 603 53 Bmo
 czech-republic
 Phone: +420543133411

OVERVIEW

Head:
 Collection Type:
 Collection Size: 10.000 - 100.000 Samples
 Gender: Female , Male

Collection storage temperature:
 Collection Description:

MATERIAL TYPE

DNA, Serum, Blood, Tissue Frozen, Tissue FFPE

AVAILABLE DATA

Preview of Directory 2.0

- ▶ Expected public release in next BBMRI-ERIC Newsletter (scheduled December 18, 2015).
- ▶ Ongoing data curation and updates.

Toward Directory 3.0/4.0/...

- ▶ Support for linked data and semantics
 - modularized data model of MIABIS 2.0
 - clarifying definition of terms
 - problem with “minimum datasets”:
what is really mandatory?
 - non-IT experts have sometimes a bit flexible view of word “mandatory”
 - MIABIS 2.0 to be ontologized
 - MIABIS (1.0) has undergone this ⇒ OMIABIS¹ & OBIB²

- ▶ Achieving interoperability though standardization: ISO TC 276 (WG1–WG5)
 - TC 276 WG5: Data Integration

¹<http://www.ontobee.org/ontology/OMIABIS>

²<http://obofoundry.org/ontology/obib.html>

Toward Directory 3.0/4.0/...

- ▶ Support for linked data and semantics
 - some ontologies are very difficult to translate: e.g., *ICD-10* ↔ *SNOMED CT transform is context dependent*
 - depends on level of detail: the more fine grained, the more difficult.
 - Support for different diagnoses ontologies in Directory 1.0
 - with syntactic search only (regex/wildcard patterns).
- ▶ Support for other data types
 - e.g., imaging biobank (working group between MIABIS and DICOM with ESR).

Toward Directory 3.0/4.0/...

- ▶ Support for capabilities of biobanks
 - increasing amount of requests to host samples/data for external projects
- ▶ Support for studies
 - collections provide 1:N mapping to samples
 - studies will support M:N mapping
 - one sample can participate in multiple studies
- ▶ Support for quality management
 - on multiple levels of aggregation,
 - starting with declaring SOP availability.

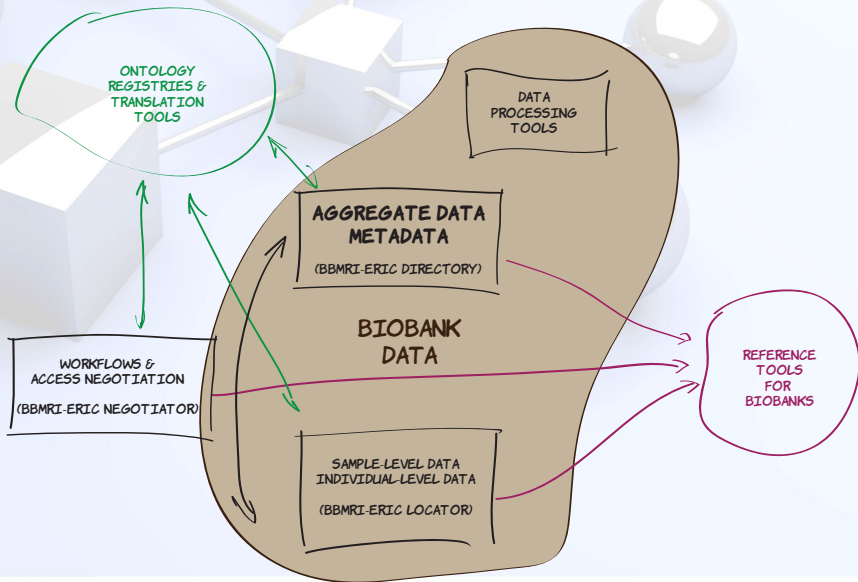
Toward Directory 3.0/4.0/...

- ▶ **BBMRI-ERIC Directory can serve as a identifier translation/mapping service:**
 - possible introduction of attributes for other identifier types for each entity.
- ▶ **BBMRI-ERIC considers also other types of identifiers (under discussion):**
 - DOI: assignment to biobanks/collections, plus possible dynamic assignment to sample sets once used for publication,
 - ongoing discussion with DataCite for BBMRI-ERIC to become DOI assigner.
 - Marker papers for biobanks/collections,
 - ORCID for affiliated people,
 - identifiers assigned by other communities.

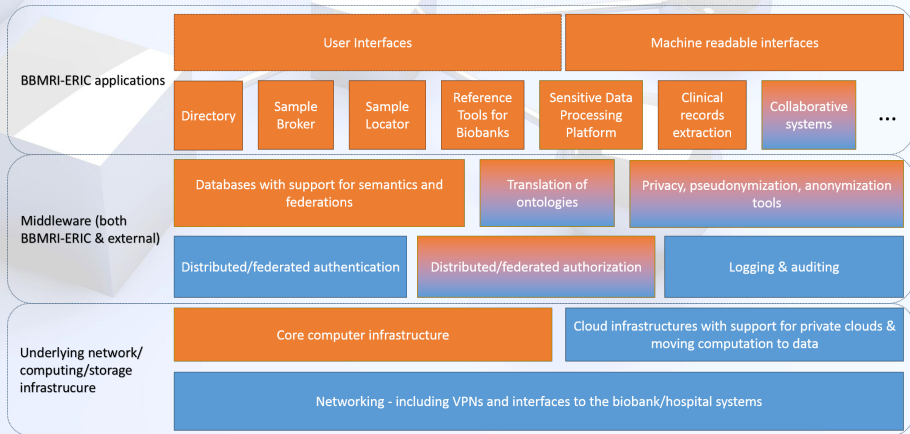
Toward Directory 3.0/4.0/...


- ▶ Metadata services and Directory become WP1 of CS IT
- ▶ Moving Directory onto the Molgenis platform
- ▶ Further development led by Morris Swertz (BBMRI.nl)
- ▶ Integration of BBMRI-LPC data
 - at the end of BBMRI-LPC project.

Toward Other Services of BBMRI-ERIC



Toward Other Services of BBMRI-ERIC





Thank you for your attention!
Q?/A!

<http://www.bbmri-eric.eu/>
<petr.holub@bbmri-eric.eu>