Metadata at BBMRI

Jan-Eric Litton

Roxana Merino

2013-03-11

BBMRI.se
BioBanking and
Molecular Resource
Infrastructure of
Sweden

About biobanks…

- Repositories of biospecimen (human) for use in research
- Different: types of biobanks, services, processes, quality controls, regulations (consent), data management (LIMS)
- Knowledge discovery about human diseases
- Towards personalized (precision) medicine

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

BBMRI.EU – Overview

- First research infrastructure funded by EC
- Main aim: Creation of a distributed bio-research infrastructure and network of biobanks with operational units in the members states
- Preparatory phase: February 2008 – January 2011 (~5 M€)
- 54 partners from 33 countries, largest infrastructure
- 7 WPs
- WP5 – Database harmonization and IT-infrastructure
  - Officially – 11 partners from 8 countries
- Continuation: BBMRI-ERIC to start in the second half of 2013
  - BBMRI-ERIC Inauguration Conference Sept.16-17, 2013 in Graz/Austria

BBMRI.EU – Data collection

1) Biobank questionnaires
2) WP5 metadata model
3) WP5 minimum dataset

1) BBMRI.EU questionnaires

1. Core (13 pages)
2. Funding (1 page)
3. ELSI (4 pages)
4. IT (10 pages)
5. Sample collections (4 pages)

1) BBMRI.EU questionnaires

- www.bbmriportal.eu
  - •315 biobanks registered
  - •20 704 864 samples (DNA, boold, serum, tissue, cell lines, etc.)

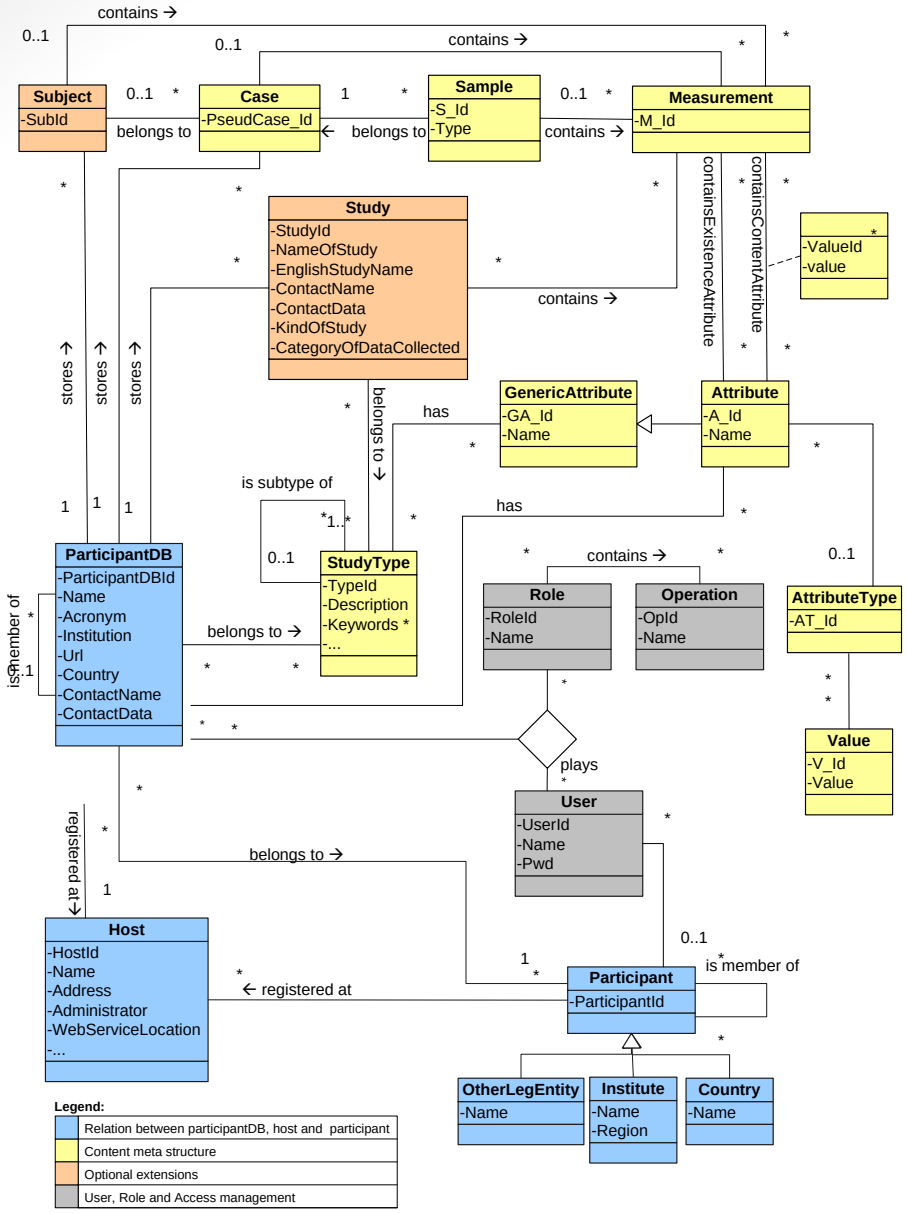
Network	38
Core	325
Collection	669
Costs and Funding	137
Resources & Methods	98
Legal, Ethics and Governance	96
Biobank IT-solutions	89
Outcome of research using biological resources	309
Detailed description of biological samples	80

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

2) BBMRI.EU WP5 metadata model

- Hub-and-spokes
- Metadata model for(national) biobank hubs

- **Needs further development**
- **Too complex?**



3) BBMRI.EU WP5 minimum dataset

Data describing biobanks

Definition	Allowed values	Explanation
BiobankAcronym	ASCII	
NameOfBiobank	Free text in English	
Institution	Free text in English	
URL		
Country	ISO-standard (3166 alpha2), two letter code	
ContactName	Free text in English	
ContactData	Free text in English	Address, Phone (E.164, No. 905 – 1.IV.2008), e.g., +46 8 524 877 59, Mail

Data describing studies

Definition	Allowed values	Explanation
NameOfStudy	Free text in any language	
EnglishStudyName	Free text in English	Translation of study name in English
ContactName	Free text in English	
ContactData	Free text in English	Address, Phone (E.164, No. 905 – 1.IV.2008), e.g., +46 8 524 877 59, Mail
KindOfStudy	Population-based, specific-disease, broad-spectrum of diseases	If "specific-disease", note ICD10
CategoriesOfDataCollected	[ClinicalDataAvailable, Diagnosis, Health information, Physiological/biochmical measures, Sociodemographic char., Socioeconomic char., Life habits/Behav., Physical environment]	Can be several values

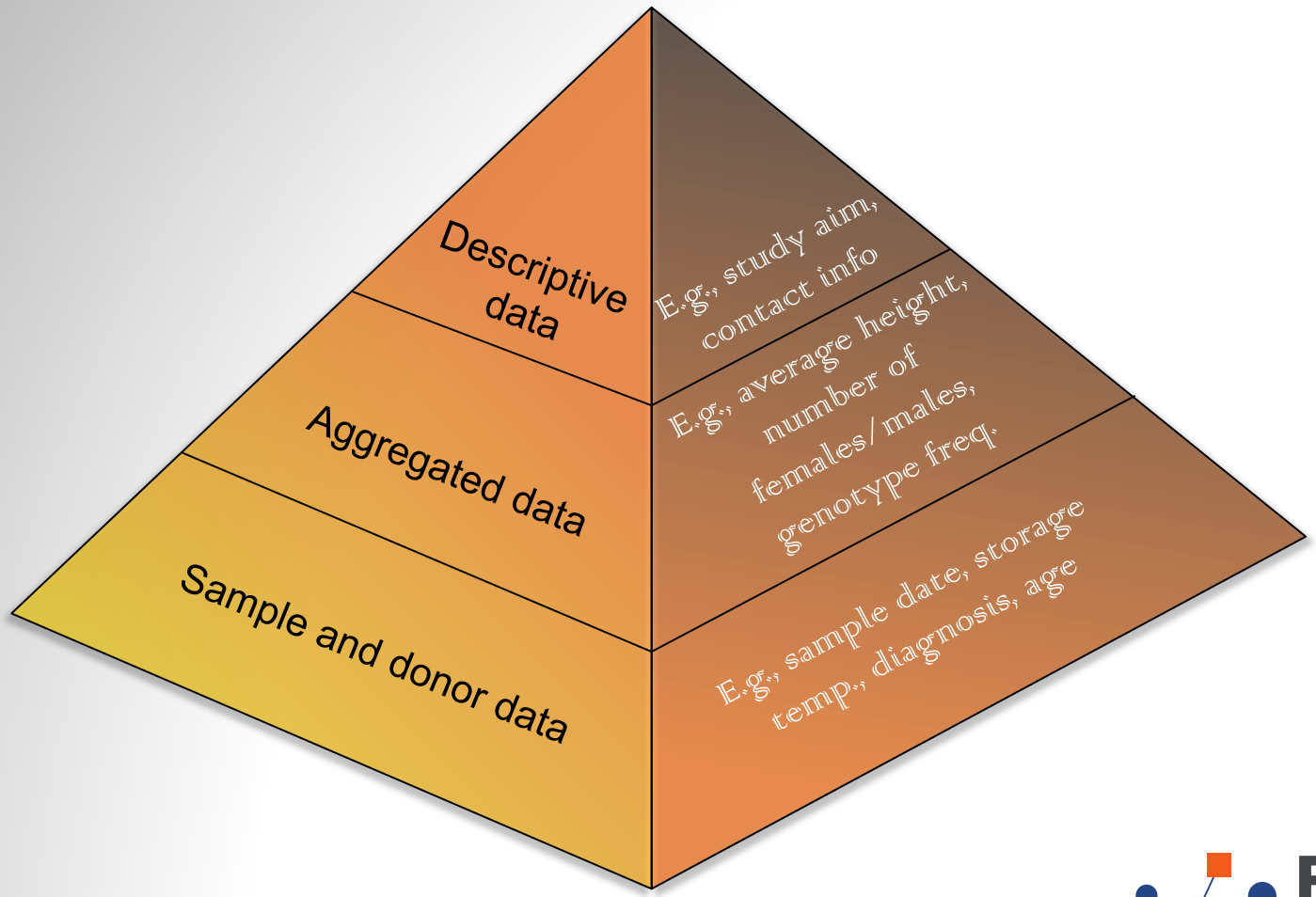
Data describing subjects/cases/samples within biobanks

Definition	Allowed values	Explanation
AgeGroup	Interval [a,b], a>0, b<200, b>=a	a and b should be selected so that k-anonymity is guaranteed. Age group of donor at time for sample collection, number of age groups determined by biobank
Gender	Male, Female, Other	Gender of subject
SampleType	DNA, cDNA/RNA, whole blood, blood cells isolates, serum, plasma, fluids, tissues cryo, tissues paraffin-imbedded, cell-lines	Type of sample. From the BBMRI core question.
SampleDate	ISO-standard (8601) time format	Date when sample was harvested
ClinicalDataAvailable	Yes/No	There exists clinical data related to the sample
OrganCategory	From the BBMRI Detailed descr bio samples	
OmicsDataAvailable	Yes/No	Genomics, proteomics etc
RestrictionsOnSampleUse	None, Consent participant, IRB approval, Approval of owner of collection	Can be several values

NOTES:

Time stamp and version control are part of the metadata schema and upload services

Data pyramid for biobank information



BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

BBMRI Nordic

- Nordic initiative BioBanking and Molecular Research Infrastructure (BBMRI Nordic) funded by the Nordic Research Council since 2010
- Collaborative network between national biobanking infrastructures in the Nordic countries
- BBMRI.se in Sweden, BBMRI.fi in Finland, BBMRI.no in Norway, Biobank Denmark, researchers from Iceland, Estonia and Faroe-Islands
- Biobank catalogs: first step towards biobank data sharing among the countries member of the network

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

BBMRI.se

- Similar structure as BBMRI.EU + “Biobank Technology” WP6
- 21 M€ from the Swedish Research Council and the medical universities
- Sweden was leading WP5 during BBMRI.eu preparatory phase

WP5 specific activities:

- *Data discovery*
- *Data integration*
- *Standards*

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Questionnaire -minimum dataset for biobanks (MIABIS)

Enkät – Register över forskningsprovsamlingar_v3

BBMRI.se Biobanking and Molecular Resource Infrastructure of Sweden

Datum för ifyllande av enkät:

Övergripande uppgifter om provsamlingen/studien

Provsamlingens ID el. studiens namn på lokalt språk (kort akronym el. dylikt):	
Provsamlingens ID el. studiens namn på engelska (även om samma som ovan):	
Kort beskrivning av provsamlingen/studiens syfte (max 120 ord):	
Provsamlingen består av (endast ett alternativ):	☐ Uttag från vård & behandlingsprovsamling ☐ Uttag från annan forskningsprovsamling ☐ Prov tagna specifikt för forskning för denna provsamling/studie
Provsamlingen startade (månad/år):	
Provsamlingen slutade alt. planeras avslutad (månad/år alt. tillsvidare):	
Destruktionsdatum (om ingen, ange ingen):	
Typ av studiedesign i den mån som detta kan anges (flera val möjliga):	☐ Fall-kontroll ☐ Kohort ☐ Tvärsnittsstudie ☐ Longitudinell ☐ Tvillingstudie ☐ Kvalitetssäkring ☐ Populationsbaserad ☐ Sjukdoms specifik ☐ Annan:
Vid sjukdomsspecifik provsamling, ange ICD-10 koder. Då detta inte är möjligt, ange diagnosen i text (för möjlighet till översättning till ICD-10 kod). Primär sjukdom: Finns data om komorbiditet:	☐ Ja ☐ Nej
Finns -omicsdata (genomics, proteomics, transcriptomics, metabolomics etc.)?	☐ Ja, finns ☐ Ja, planeras ☐ Nej

Kontakt Loreana Norlin Tel: 08-524 87408 E-post: loreana.norlin@ki.se
Kontakt Sonja Eaker Tel: 018-15 19 20 E-post: sonja.eaker@surgsciuu.se
Adress BBMRI.se Nobels väg 12A 171 77 Stockholm

Enkät – Register över forskningsprovsamlingar_v3

BBMRI.se Biobanking and Molecular Resource Infrastructure of Sweden

Om ja, ange vilken -omicsdata.	☐ Genomics ☐ Proteomics ☐ Transcriptomics ☐ Metabolomics
Om ja, ange metod (helgenom, Illumina etc.):	

Uppgifter om provgivarna

Då provsamlingen ej är påbörjad eller pågående ifylles planerad information (enligt etikansökan eller motsvarande).	
Kön (flera val möjliga):	☐ Kvinnor ☐ Män
Antal individer i provsamlingen:	
Åldersgrupp vid provtagning (yngsta och äldsta personen som ingår i provsamlingen):	

Uppgifter om prov

Då provsamlingen ej är påbörjad eller pågående ifylles planerad information (enligt etikansökan eller motsvarande).	
Vilken typ av biologiskt material inhämtas från individerna (flera val möjliga):	☐ Helblod ☐ Plasma ☐ Serum ☐ Urin ☐ Saliv ☐ CSV ☐ DNA ☐ RNA ☐ Faeces ☐ Vävnad (ange typ nedan) ☐ Celler (ange typ nedan) ☐ Annat, specificera: Typ av vävnad eller celler:
Temperatur för långtidsförvaring (flera val möjliga):	☐ Rumstemperatur ☐ -35°C till -18 °C ☐ -60 °C till - 85°C ☐ Flytande kväve ☐ +4 °C ☐ Annat, ange:
Sociststyrelsens reg.nr på biobank där provsamlingen ingår:	

Kontakt Loreana Norlin Tel: 08-524 87408 E-post: loreana.norlin@ki.se
Kontakt Sonja Eaker Tel: 018-15 19 20 E-post: sonja.eaker@surgsciuu.se
Adress BBMRI.se Nobels väg 12A 171 77 Stockholm

Enkät – Register över forskningsprovsamlingar_v3

BBMRI.se Biobanking and Molecular Resource Infrastructure of Sweden

Övriga uppgifter om data förknippad med provsamlingen/studien

Vilken typ av information utöver biologiskt material inhämtas om provgivarna (flera val möjliga)?	☐ Registerdata ☐ Enkätdata ☐ Fysiska mätningar ☐ Bilddata ☐ Annat, specificera:
Vilka av följande ämnesområden innefattar datat ovan (flera val möjliga)?	☐ Socioekonomi/demografi ☐ Upplevt hälsotillstånd ☐ Levnadsvanor/livsstil ☐ Mediciner ☐ Kvinnohälsa ☐ Reproduktiv hälsa ☐ Individuell hälsohistoria ☐ Fysisk miljö ☐ Skador ☐ Familjemedlemmars hälsohistoria ☐ Mental hälsa ☐ Annat
Övrig information som rör provsamlingen eller studien (max 250 ord).	

Kontaktuppgifter för frågor gällande provsamlingen/studien

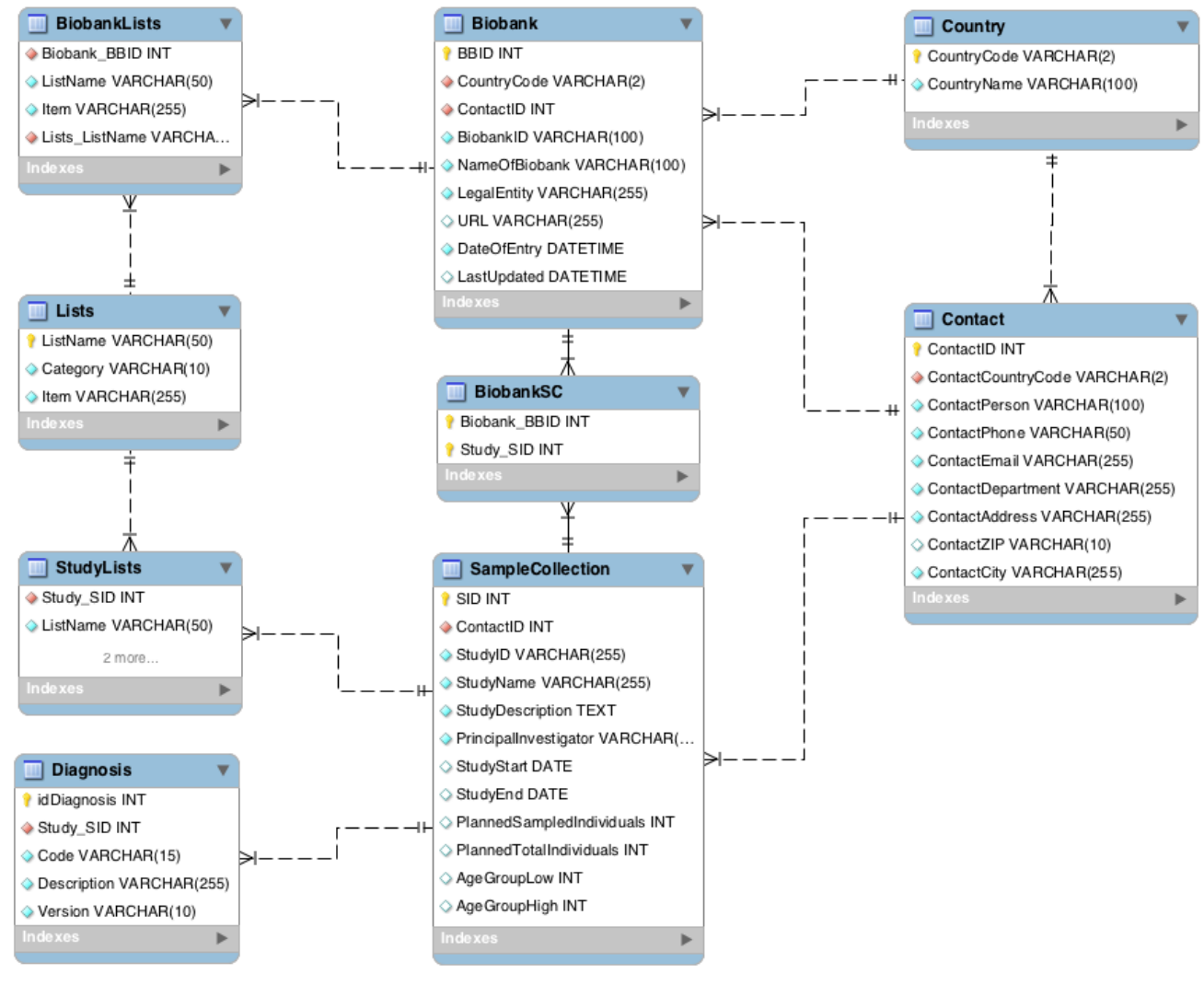
Provsamlingsansvarig eller Principal Investigator:	
Kontaktperson:	Telefon:
E-post, kontakt:	
Adress, kontakt (inkl. institution, avdelning etc.):	

Tack för din medverkan!
Vänligen återsänd enkäten ifylld till loreana.norlin@ki.se

Kontakt Loreana Norlin Tel: 08-524 87408 E-post: loreana.norlin@ki.se
Kontakt Sonja Eaker Tel: 018-15 19 20 E-post: sonja.eaker@surgsciuu.se
Adress BBMRI.se Nobels väg 12A 171 77 Stockholm

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Minimum Information About Biobank data Sharing (MIABIS)



Norlin et al.
A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS. *Biopreservation and Biobanking*. August 2012, 10(4): 343-348.

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

BBMRIRegister

bbmriregister.meb.ki.se:8080/AwareIM/logonGuest.aw?domain=BBMRIRegister

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

BBMRI.se
Sample Collection Register

Contact us | Login | Log Out

- Home
- Find Sample Collections
- Help

External Links

- BBMRI.se Wiki
- Catalog of European Biobanks (User: guest, Password: catalogue)
- KI Biobank studiekatalog
- Svensk Nationell Datatjänst (SND)
- Molecular Methods Database
- Tubafrost Central Database
- NordiCDB
- Deutches Biobanken-Register
- EuroBioBank
- NCI Biospecimen Research Database
- Danish National Biobank
- P3G

Welcome | Organizations | Sample Collections

19/11/2012

Welcome to the BBMRI.se Register!

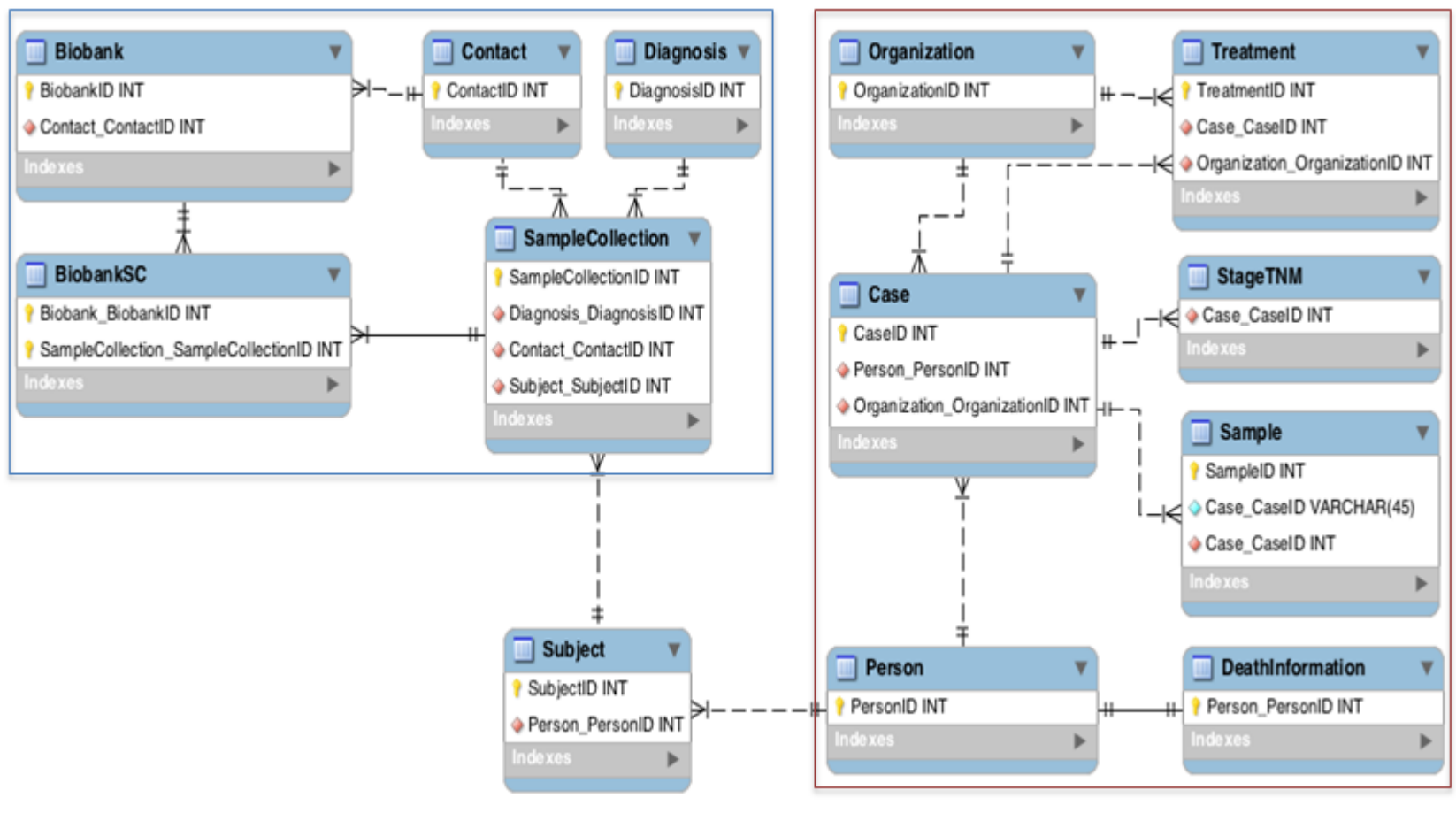
This website is intended to help researchers initiate collaborations. Here you can find information about sample collections containing human biological material as well as survey and register data associated with the donors. For further information, contact the responsible people for each sample collection.

Total number of donors: 1150855

MIABIS - Eurocourse

MIABIS

Eurocourse



BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Some comments…

- We are aiming to have a stable data model for biobank data sharing (MIABIS, omiabis)
- Descriptive metadata about omics data from biobanked samples hasn't been formalized
- Important efforts on data sharing (GEO, EBI-EGA)
- Lot of ontologies! (BioPortal)
- Interesting project for bio-resources identification (PID) (BRIF) (eagle-i) (ORCID)
- EU's new Personal data protection legislation is a big issue
- Informatics integration of biobanking, research and clinics is still in an early phase

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Big data and Biobanking

- The volume of data produced by genomics research (e.g. NGS) is increasing at a higher rate than Moore's Law predicts
  - First human genome took a decade to complete and 3 years for data analysis at the cost of **$3 billions**.
  - An entire human genome has **3.3 billions of base pairs**. It can be sequenced and analyzed in a few hours for a few thousand dollars.
  - Base pairs are read in short sequences and then assembled. It is done multiple times to produce an accurate sequence (raw file format, up to 30 terabytes).
  - A proteomics experiment can create more rows of data than a traditional row-based DB can handle
- Omics data formats vary depending on the technic and software (e.g. open-XML data format (proteomics), SFF, Fasta, fastq (NGS))
- Storage and analysis of omics data require a lot of store capacity and computing power
- An ongoing EU project will be a model for omics data storage and analysis (**BiobankCloud**: http://www.biobankcloud.com/)
- e-Science for Cancer Prevention and Cure eCPC

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Big data and Biobanking: Sweden

- BBMRI.se is working towards a secure long term storage and analysis of big data from research studies on human samples
- A register over studies in Swedish bio-medical research institutions was launched in November 2012:
  - BBMRI.se Sample Collection Register
  - To date, over 80 studies have been registered from which ~32 are conducting omics experiments on ~276074 donors. If just a single sequence of each donor will be stored, ~276074 * 30 terabytes of storage will be needed (this a very small fraction of the real amount of data)
  - Another major issue is the omics data analysis
  - Up to date, researchers find their own solutions for omics data storage and analysis

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Big data and Biobanking: Big issues

- The omics data resulting from research on biobanked samples is expected to increase dramatically over the next years. Some important matters need to be taken into consideration:
  - Secure long term storage capability (versioning genomes, etc.)
  - Tracking of omics data use and reuse (analysis results)
  - Fragmentation of knowledge among different omics
  - Standardization of omics data representation for data sharing
  - Standardization of omics data representation for analysis and interpretation
  - Personal data privacy protection
    - Some omics data can lead to the identification of the sample donor

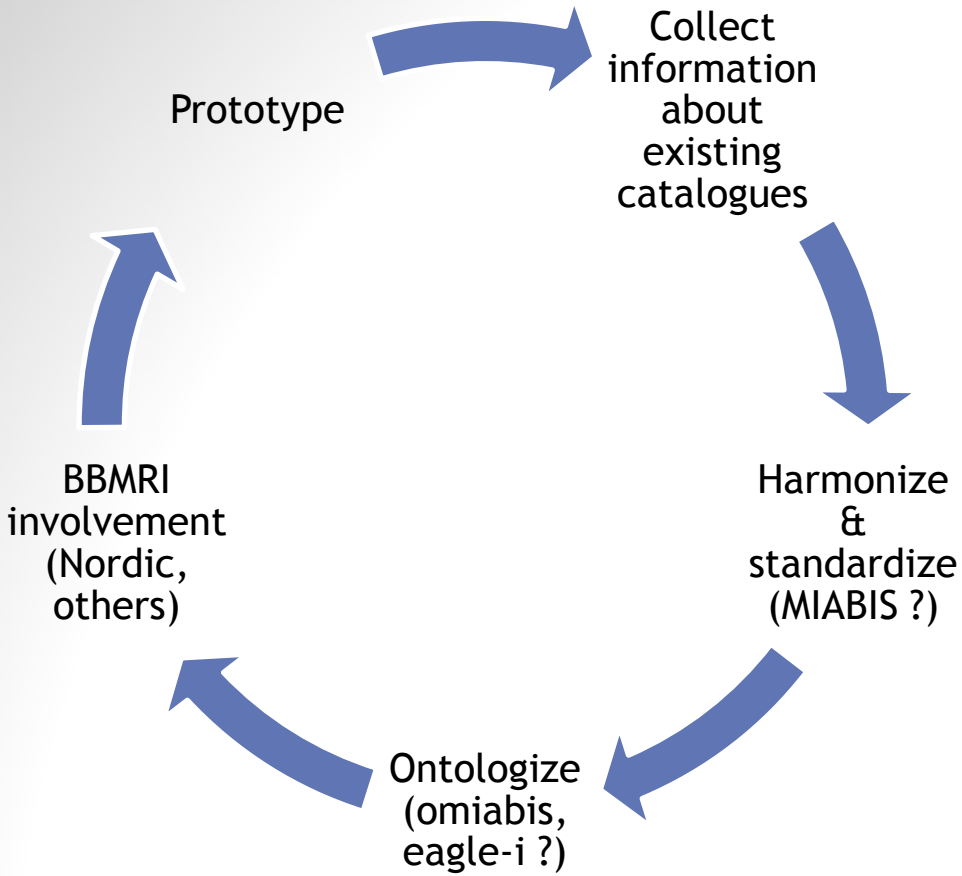
BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Can EUDAT help to build a catalogue for BBMRI?

- Main user: Researcher
- Main aim: Search for bio-resources availability
  - How deep to go? (any bio-resource, sample, sample data)
- Sample data at the metadata and aggregated level
- No personal data protection issues
- Centralized catalogue?
  - Biobanks, researchers, vendors… upload data (complicated)
- Distributed catalogue?
  - National catalogues upload data to EUDAT catalogue

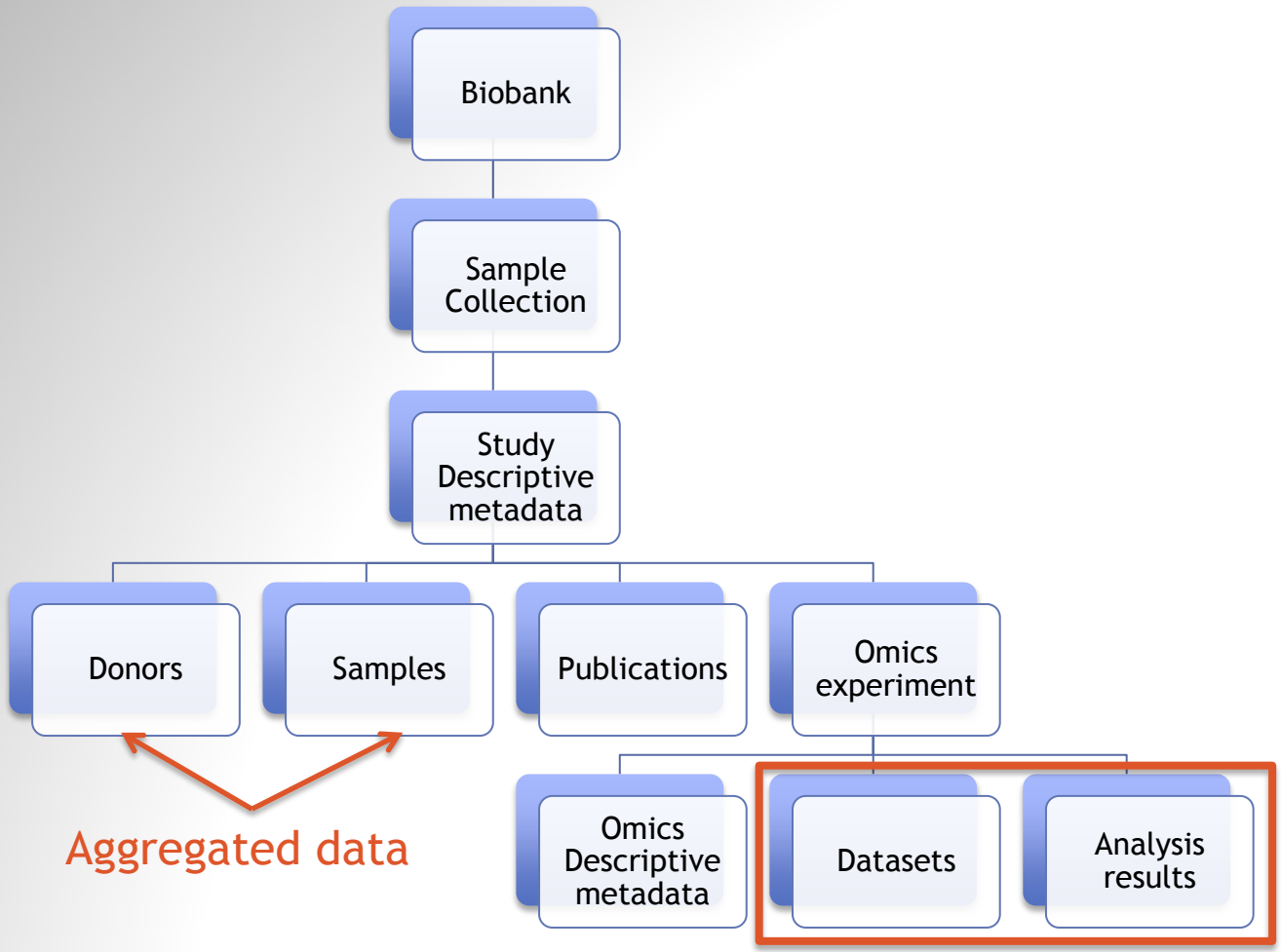
BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

EUDAT Catalogue for BBMRI?



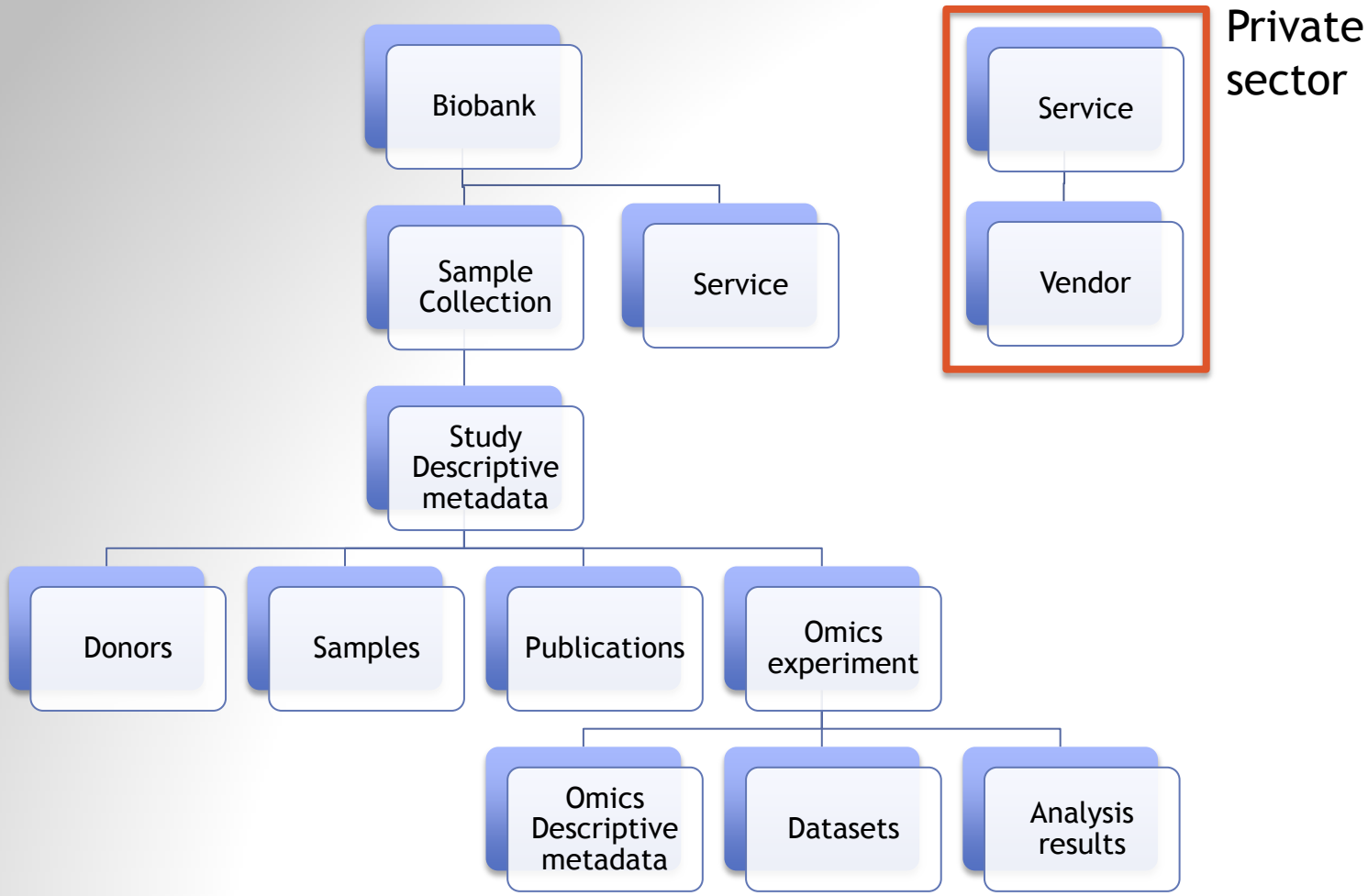
BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Search for sample & sample data



BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Search for any Bio-resource



BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Conclusions

- Projects as EUDAT would be of great benefit for BBMRI
  - storage infrastructure for omics data
  - Relevant information about sample and sample data availability all over Europe
  - Easy to use (dropbox, eage-i)
- BBMRI does not have a stable metadata model for biobanking but there is sufficient work done in that direction
- Networking biobanks and biomedical research institutions - big challenge
- Implementation of a BBMRI catalogue should be divided in several stages, from the general information about biobanks down to the specific data about samples and data analysis
- Interdisciplinary catalogue: social statistics, bio-statistics, modeling, environment, ethics and law, medical terms (SNOMED, ICD codes, etc.)
- Omics data storage and analysis is a bottleneck in biomedical research due to the sizes and the diversity of data formats and analysis methods

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden

Thanks!

Tack!

Gracias!

Jan-Eric.Litton@ki.se
Roxana.Martinez@ki.se

BBMRI.se
BioBanking and Molecular Resource Infrastructure of Sweden