EuSoMII ACADEMY

Game changers in radiology

The impact of informatics and A.I. on medical imaging

Jointly organized by EuSoMII, SIIM and Erasmus MC

Rotterdam

November 18th, 2017

Image sharing and biobanks

Bernard Gibaud

MediCIS, LTSI, U1099 Inserm
Faculté de médecine, Rennes
bernard.gibaud@univ-rennes1.fr

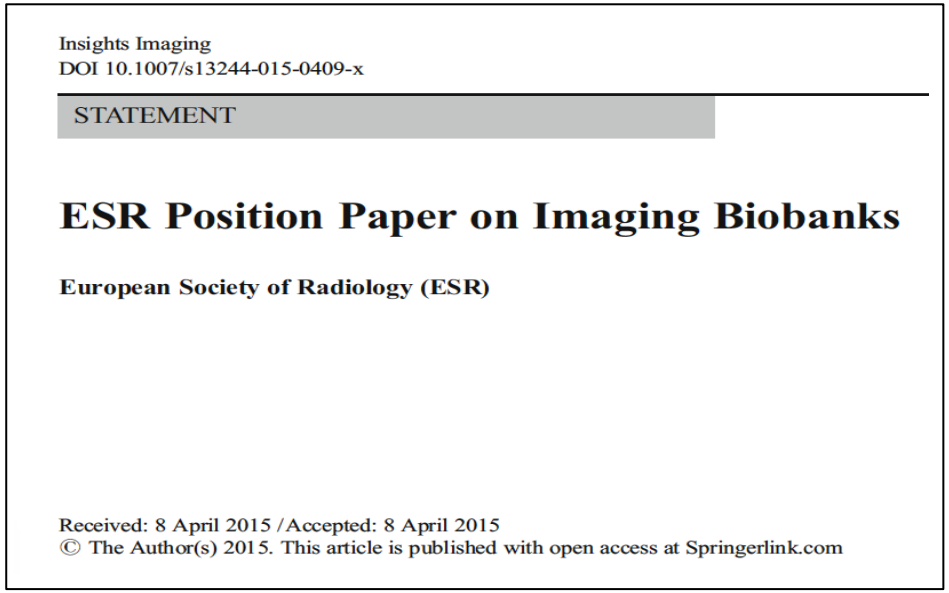
Overview

- Introduction
  - Origin of the activity
  - Motivations
- Image biobanks and biobanks
  - Analysis of the workflow
  - Need for federations
- From concept to implementation
  - Collaboration ESR / BBMRI-ERIC
- Discussion

Origin of the concept

- 2014: Working group on **Imaging Biobanks** (Chair: E. Neri)
- Extracts from the ESR Mission statement
  - « to promote the development of imaging biobanks and intelligent tools for the analysis and processing of biomarkers »
  - « to stimulate the link between imaging biobanks and traditional biobanks through the development of standards »
  - « to monitor the existing imaging biobanks in Europe, to promote the federation of such imaging biobanks, and to elaborate a white paper on imaging biobanks»

Imaging biobanks



Insights Imaging
DOI 10.1007/s13244-015-0409-x

STATEMENT

ESR Position Paper on Imaging Biobanks

European Society of Radiology (ESR)

Received: 8 April 2015 / Accepted: 8 April 2015

2015: Paper prepared by the ESR Working Group on Imaging Biobanks

Motivations

- Personalised medicine context
  - Take into account all available information to take the best decisions for patient management (images, genomic, other biological information)
  - Growing role of computerized systems to determine
    - What are the best decisions, learned from past experience (deep learning)
    - How to assist physicians in decision making

      → through formal decision models
- Key importance of imaging and imaging biomarkers
  - in all kinds of decisions (diagnostic, therapeutic, pronostic, prediction of effect of therapy, assessment of effect of therapy, etc.)

Big data context

- New opportunities arising from the Big data context
  - to re-use data for other purposes (not anticipated)
  - to learn from huge quantities of data (deep learning)
- New opportunities for biomedical research
  - Re-classify the (retrospective) data based on new knowledge
  - Suggest and test (in silico) new decision models
    - (which need to be tested in prospective studies)
  - Tremendous possibilities for imaging (because imaging is data)

Challenges

for personalised medicine in a big data context

- Ability to gather (i.e. to store and communicate) data from multiple clinical/research sites
  - To pool as many cases as needed to get the relevant statistical power
  - For each case, to gather the various information needed to answer each research question
    - imaging / biological data / genomic data / clinical data
  - To access specimen when new analyses are needed
- Non-technical challenges
  - Protection of humans: regulation on data sharing in research in humans
  - Confidentiality risks: legal barriers
  - Reluctance of researchers w.r.t. data sharing
- Technical challenges
  - To provide a technical infrastructure to support such information storage and exchange
  - Ensuring semantical consistency within pooled datasets
    - Availability of standards for describing shared entities

Imaging biobank

- Definition:
  - « organised databases of medical images, and associated imaging biomarkers (radiology and beyond), shared among multiple researchers, linked to other repositories » [Mission statement of ESR WG on imaging biobanks]

Imaging data
(definition)

- Acquired images (i.e. acquired on humans, animals, or specimen from these organisms)
- Derived data, generated through automated data processing
  - Images, e.g. denoised images, corrected of bias distorsion, template-resampled images, statistical maps, etc.
  - ROI, e.g. binary masks (segmentation results), graphs (e.g. contours, tractography, 3D surfaces 3D as meshes)
  - Measurements made from image data (imaging biomarkers)
- Metadata associated to these data …

Metadata associated to imaging data

- Descriptive Metadata
  - Identifier of the dataset
  - Image characteristics (number of lines, columns, etc)
  - Image type (e.g. ref. to an image taxonomy)
  - Format descriptor
  - Link to a container or access resource
- Provenance Metadata
  - For acquired images : acquisition protocol and parameters, contrast agent, activation paradigm, radiation dose etc.
  - For derived data: nature of processing, processing tools, inputs, outputs, parameters
- Contextual Metadata
  - Study in the context of which the dataset or data item was created
  - Relationship to other data related to the same subject, potentially relavant to the research question(s) at stake
    - Clinical data (surgery, radiotherapy, etc)
    - Neuropsy. or behavioural tests
    - Genomics and other omics

Biobank

- Biobank
  - « Repositories of resources for medicine and medical research
    - Biosamples
    - Accompanying data: clinical, phenotypes, lifestyle, …
    - Data generated from samples: imaging, omics, …
    - Expertise and services: sample and data hosting, … »

[Holub 2016, derived from definitions from Fransson 2015]

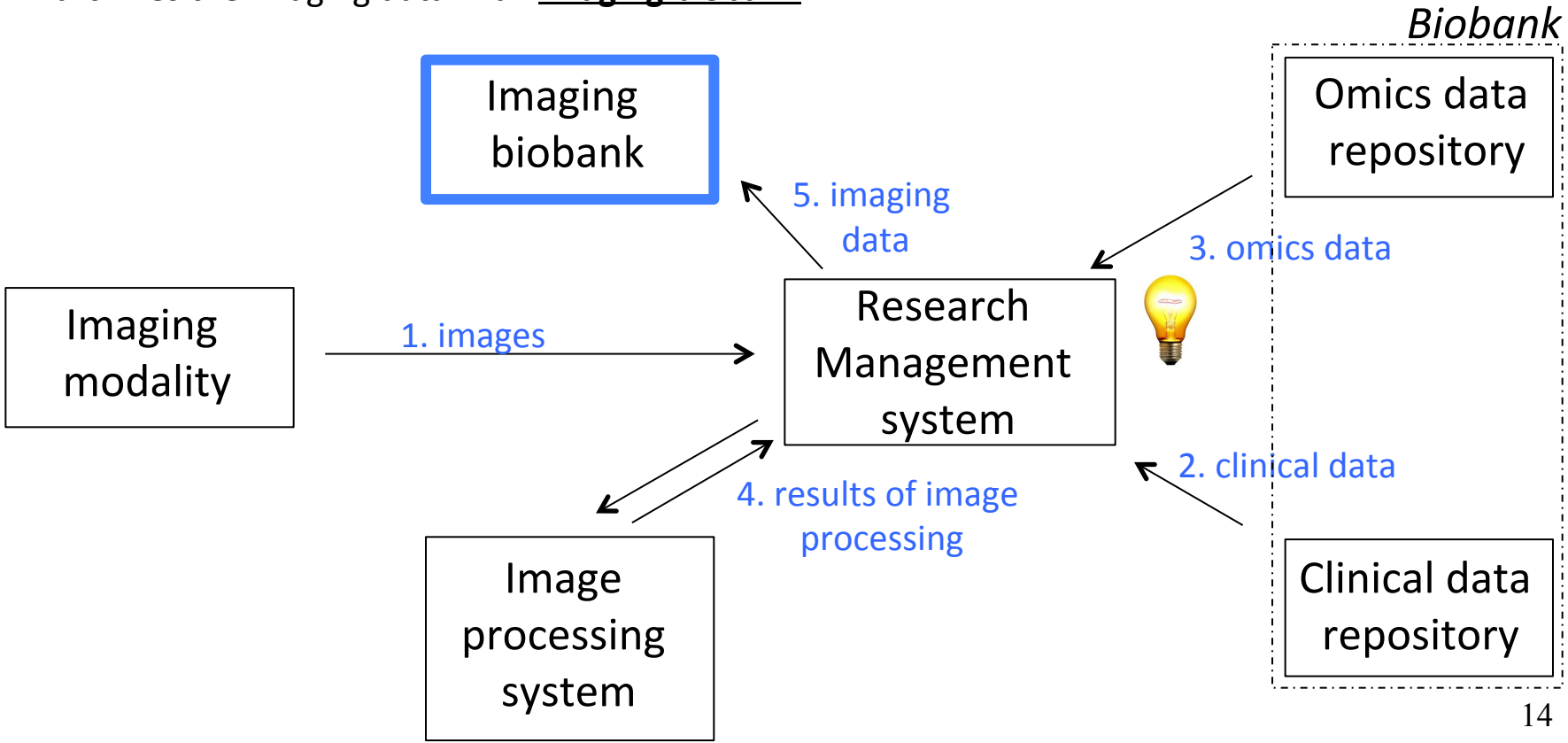
Workflow between Imaging biobanks and biobanks

- **First scenario**: Imaging biobanks as a resource to share and reuse imaging data, after a primary use in a research study

- **Second scenario**: Imaging biobanks as a resource to share and reuse imaging data, before any use in research studies

First Scenario: a Research Management System

- collects the data
- supports data processing and answering the research question
- archives the imaging data in an **imaging biobank**



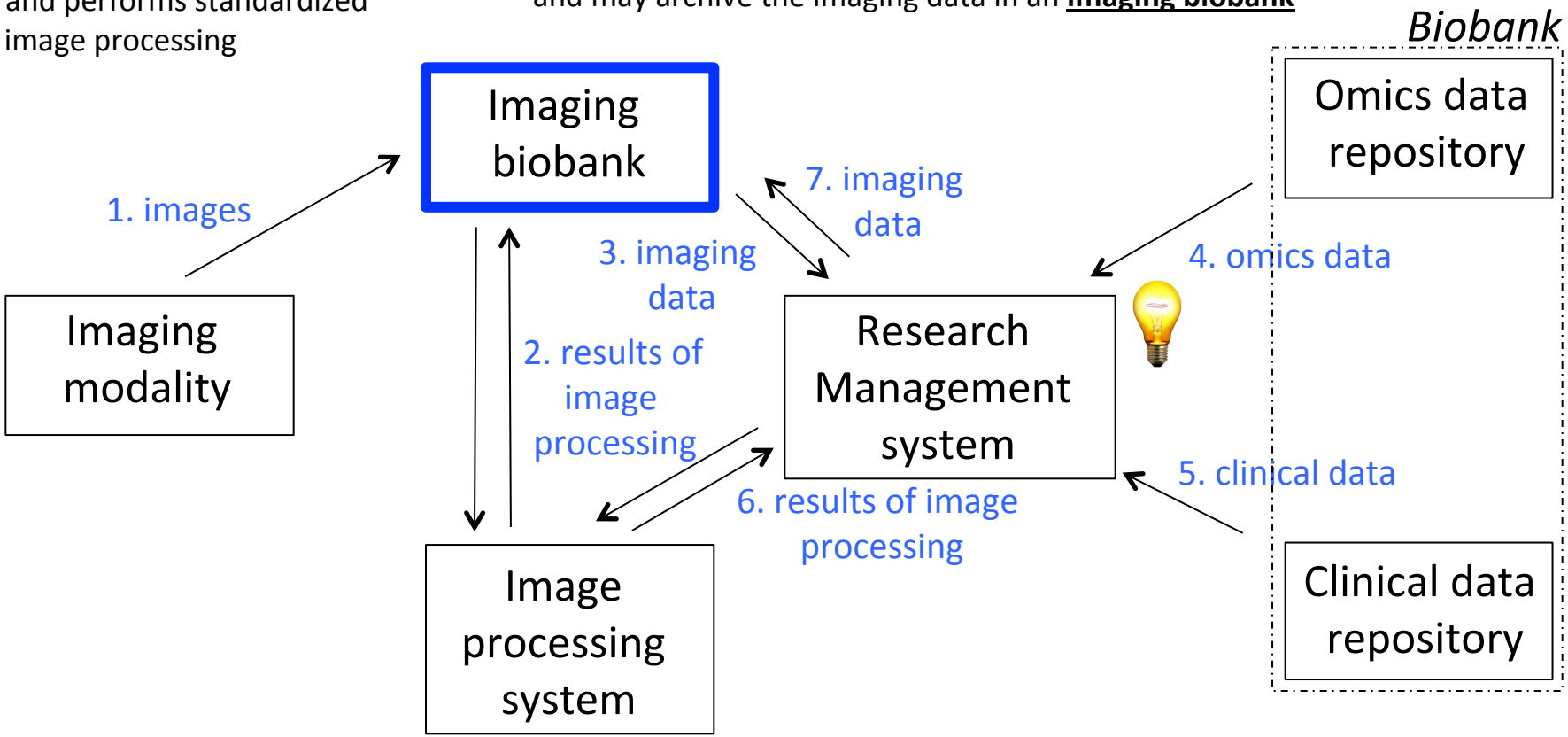
Second scenario:

the Imaging biobank

- collects data
- and performs standardized image processing

Then, some **Research Management System**

- collects the data
- supports data processing and answering the research question
- and may archive the imaging data in an **imaging biobank**



Biobanks & Imaging biobanks

Biobanks

- Specimen
- Data about subjects and studies
- Data about specimen
- Derived data
- Provenance data (specimen, derived data)

Imaging Biobanks

- Acquired images
- Derived data (images, ROIs, biomarkers)
- Descriptive metadata
- Provenance metadata (acq. data, derived data)
- Contextual metadata (data about subjects and studies)

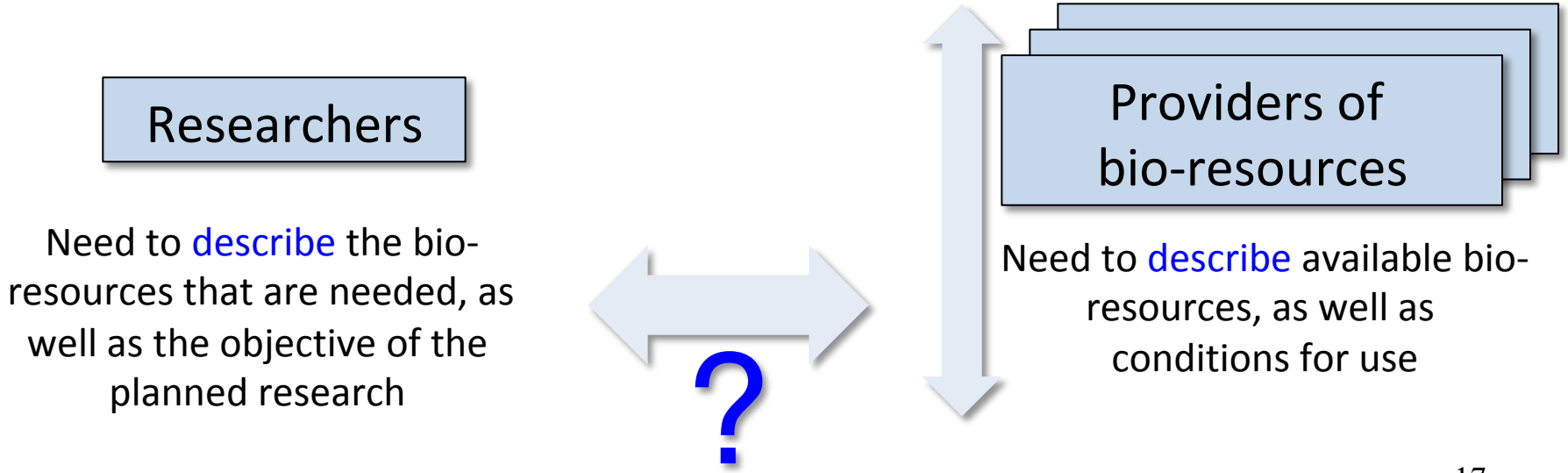
Commonalities: - both need to represent data about subjects and studies
- both need to manage derived data and related provenance data

Differences: - biobanks manage physical specimen (imaging biobanks manage only data)
- biobanks may include image data

Need for federations

Basic problem to solve

- Help researchers find the bio-resources they need for their research (e.g. biological specimen, images)



Need for federations

- Need for a common language suitable to express queries about existing bio-resources
  - Suitable to cope with a very broad domain, and able to refer to the domain entities at various abstraction levels
- Need to link data, e.g.
  - from 

  or conversely,
  - from 

→ Semantic web technologies provide such languages and tools (ontology languages, RDF-based linked data, SPARQL query language)

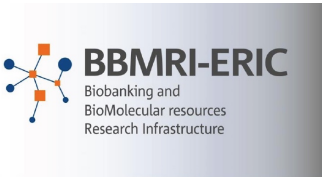
Information to model

- Bio-resources, i.e. any biological sample or information content pertaining to some biological organism*, e.g.
  - Biological specimen, e.g., tissue, blood, plasma, saliva, urine, DNA, RNA, isolated pathogen
  - Images, such as medical images (e.g. CT, MRI, NucMed, US) or pathology images
- Related phenotypic information
- Collections, and related research goals
  - Goals may be defined in general terms (e.g. population studies) or specific terms (precise study)

* This definition slightly differs from the one given in : Mabile et al. GigaScience 2013, 2:7

From concept to implementation

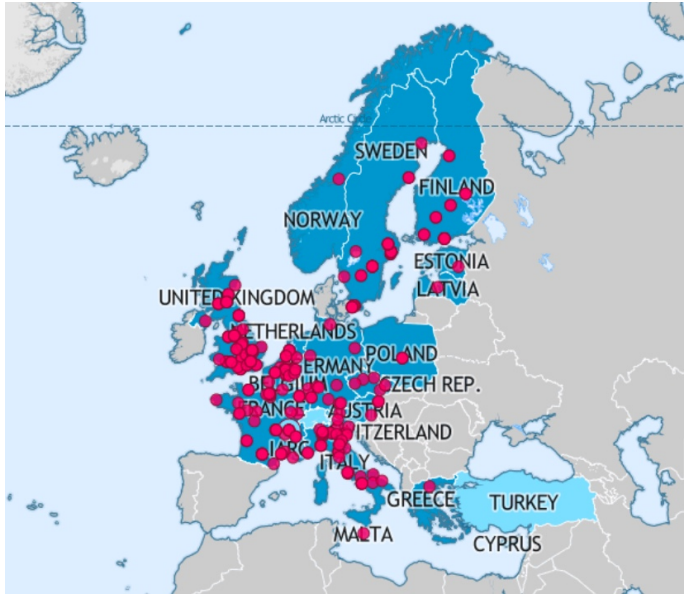
- Collaboration between 

and 

  - « BBMRI-ERIC is a distributed research infrastructure of biobanks and biomolecular resources »
- Objectives
  - **Short-term** objective: extend the MIABIS-based DIRECTORY model of BBMRI-ERIC to include image collections
  - **Medium-term** objective: collaborate to extend the OMIABIS model for imaging and map it to DICOM

BBMRI-ERIC Directory

- The BBMRI-ERIC Directory is « a tool to share aggregate information about biobanks that are willing external collaboration »
- It provides descriptions of the collections of bio-specimen
- Based on the MIABIS 2.0 data model
  - biobanks – institutional envelopes
  - collections – information on samples and data
  - biobanks networks
  - contact information



Extension of the BBMRI-ERIC Directory model

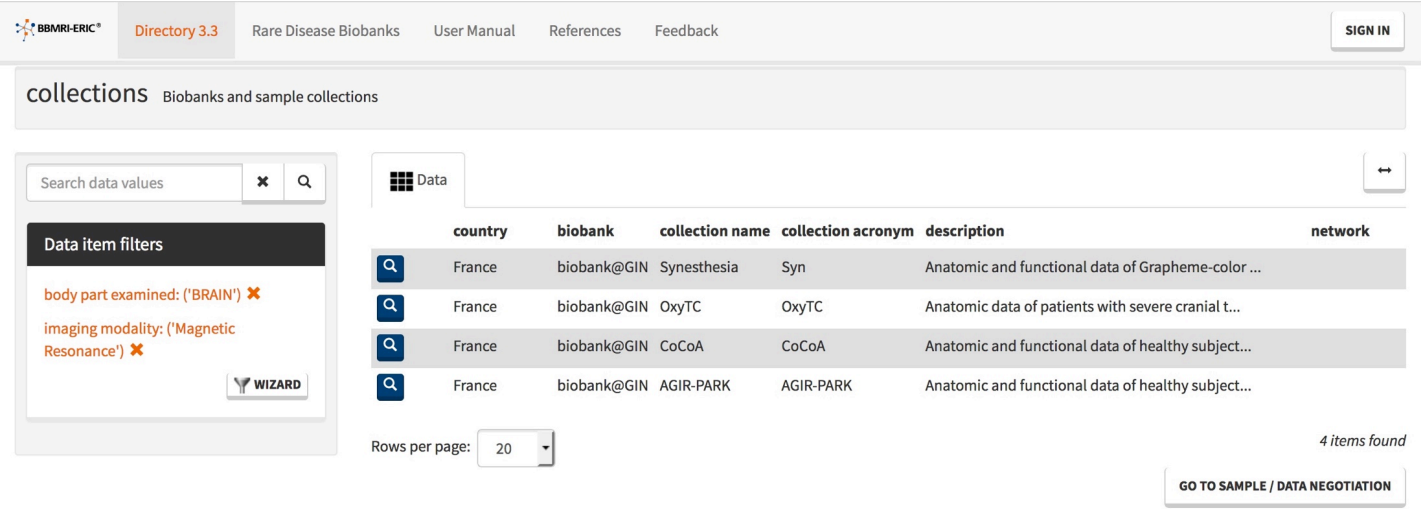
NEW: imaging collection: « collection of medical images »

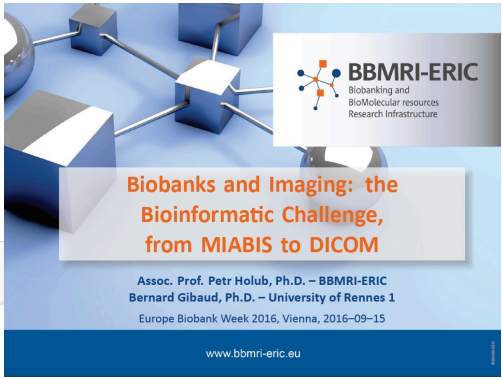
- id
- acronym
- name
- description
- body part examined (116 terms)
- imaging modality (54terms)
- image dataset type (120 terms)
- sex
- age low
- age high
- age unit
- link to other data
- collection type (cohort, cross-sec, long., etc)
- disease
- head name
- data access fee (Y/N)
- joint project (Y/N)
- access description
- number of images
- timestamp
- commercial usage allowed (Y/N)
- contact person
- citation reference

Donors (of images) described in the same way as donors (of samples)

Image collections: preliminary tests

- Description of image collections from the GIN (collab. Michel Dojat, Grenoble)





presented at the Europe Biobank Week 2016 (Holub *et al.*)

Extension of the OMIABIS Ontology and mapping with DICOM

- Two ontology models exist
  - OMIABIS (2014), Ontologized version of the MIABIS 2.0 model
  - OBIB (2016), Ontology for Biobanking (derived from the latter)

Brochhausen *et al. Journal of Biomedical Semantics* 2013, **4**:23
http://www.jbiomedsem.com/content/4/1/23

JOURNAL OF BIOMEDICAL SEMANTICS

RESEARCH Open Access

Developing a semantically rich ontology for the biobank-administration domain

Mathias Brochhausen[1*], Martin N Fransson[2], Nitin V Kanaskar[3], Mikael Eriksson[2], Roxana Merino-Martinez[2], Roger A Hall[1], Loreana Norlin[2], Sanela Kjellqvist[2], Maria Hortlund[2], Umit Topaloglu[1], William R Hogan[1] and Jan-Eric Litton[2]

Brochhausen *et al. Journal of Biomedical Semantics* (2016) 7:23
DOI 10.1186/s13326-016-0068-y

Journal of Biomedical Semantics

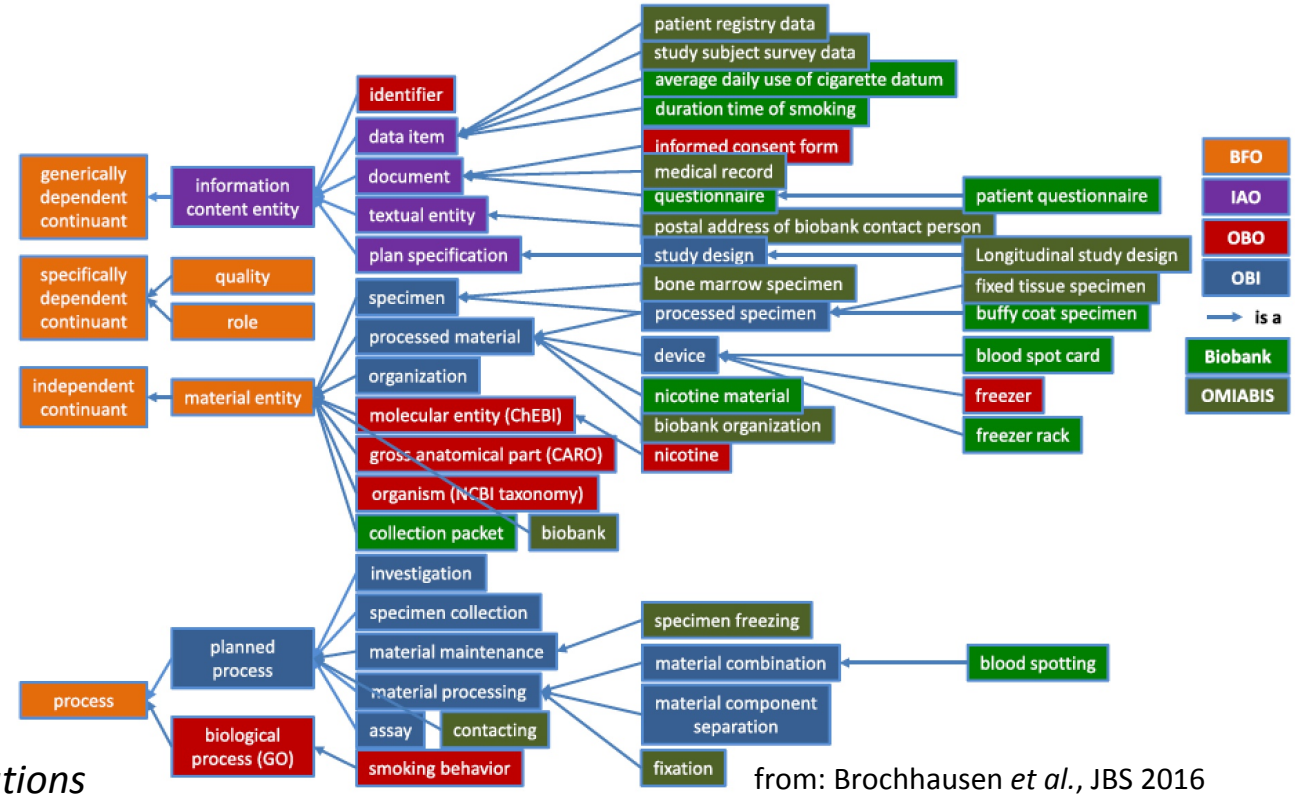
RESEARCH Open Access

OBIB-a novel ontology for biobanking

Mathias Brochhausen[1*], Jie Zheng[2], David Birtwell[3], Heather Williams[3], Anna Maria Masci[4], Helena Judge Ellis[5] and Christian J. Stoeckert Jr.[2]

Central classes of OBIB

- OBIB is based on
  - BFO[1]
  - OBI[2]
  - IAO[3]



from: Brochhausen *et al.*, JBS 2016

1 Basic Formal Ontology
2 Ontology of Biomedical Investigations
3 Ontology of Information Artefacts

Extending OMIABIS/OBIB

- Domain to be covered : Image data and metadata
  - Need to map the DICOM vocabulary
  - And to reuse as much as possible existing ontology resources
    - Upper level ontology: BFO
    - Measurement and information artifacts: OBI / IAO
    - Qualities: PATO
    - Provenance: PROV
    - Imaging: RadLex, AIM, OME
    - Imaging datasets: OntoNeuroLog
    - Medicine in general: SNOMED CT, ICD, NCIT, etc.
    - Imag. biomarkers: QIBO (Buckler *et al.*), BiomRKRS (Ofoghi *et al.*)

Discussion- BBMRI-ERIC Directory

- Description of image collections
  - Limited to acquired images, no description of processed images
  - No sub-modalities, e.g. DTI, DCE, fMRI, ASL etc.
  - Potential problems with DICOM ‘Body part examined’ (optional)
  - No reference to imaging biomarkers

Discussion- BBMRI-ERIC Directory

- Acute need for more extended testing, e.g. using image collections from
  - large population studies, especially
    - the UK biobank
    - the NAKO in Germany
  - or other imaging studies
    - from BBMRI-ERIC NL
    - from France Life Imaging repositories in France

Discussion- OMIABIS-OBIB extension

- The domain to covered is quite large
  - General taxonomy of image datasets
  - General taxomomy of image processing
  - General taxonomy of image biomarkers
  - Representation of provenance for both acquired and derived imaging data

  → Collaboration with DICOM and RadLex acutely needed
- Domain overlap between medical imaging and histopathology
  - A roadmap was discussed at an imaging workshop organised by Barry Smith in Buffalo in 2014, followed by a white paper in 2015

Conclusion –
Take home messages

- The successful use of imaging biobanks in personalised medicine research requires:
  - high precision in terms of semantics (can be provided by ontologies)
  - maximum interoperability with biobanks, for all aspects that are not specific of the medical imaging domain
  - concerning imaging, ontologies are primarily needed for image datasets, image processing, and image biomarkers.
- It is also necessary to agree on suitable data structures (DICOM, DICOM SR, and other existing formats used in research, e.g. BIDS);
- Both bottom-up (e.g. work from existing image collections, image processing packages) and top-down (more abstract definition of all entities) approaches are relevant and useful.

Thank you for your kind attention !

Acknowledgements

- Hans-Ulrich Kauczor (ESR, Heidelberg)
- Members of the ESR WG on imaging biobanks
- Members of the MIABIS-Imaging WG, especially Petr Holub (BBMRI-ERIC)
- BBMRI-ERIC Common Service IT
- Michel Dojat (GIN, Grenoble)