Step of a Biobank design and implementation

Some concerns:

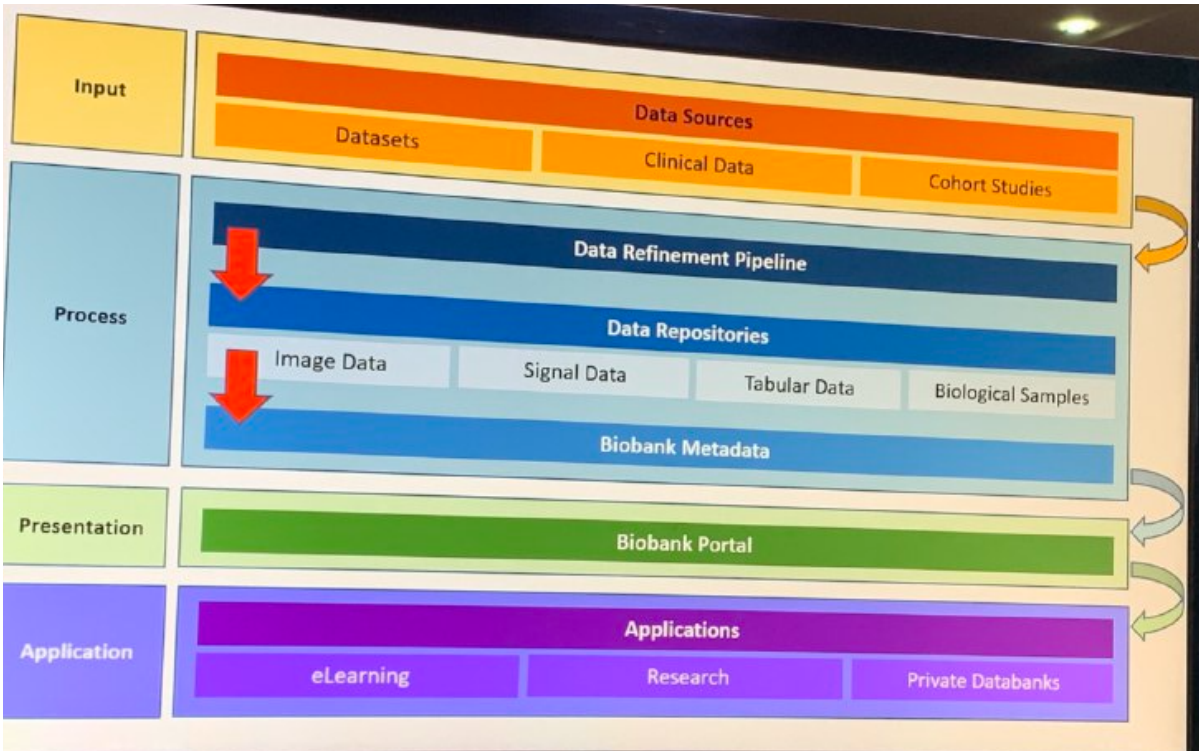
How to prevent Defacing? Should we ?

How to let others use our bank? Federated learning?

What structure is the right one? Dictionary is an answer.

What about the image quality?

How to actually query biodata in a way that is patients privacy is not violated and is most convenient for researchers?



Main parts:

Gathering information about devices that take images(MRI , ...)

Making a unique structure to input our models and make every image from different places standard.

Make sure we can accept donations form other banks

Data cleaning

Building models and perform data analysis
Building a database , app , website , ...(sofware engineering)

Planning step (the more you put your time into it the better the outcome)

Collect information about users, site needs, consent requirements, and support needs of the target population.
Gathering information about currently used imaging devices such as how thy are storing the image , patient data and quality of images.
Also study other successful biobanks.

Data:

Data pre-processing

- *cleaning*, which means using only complete data;
- *reduction*, which means that the data follow a specific model and only the data with model parameters can be used;
- *transformation* means conversion of data to the specific format for intended analysis and discretization, which divides data to special sets like subgroups, intervals, subsets, and files

Data storage(cloud storage is the best answer):

Data is stored, and it is necessary to take into consideration that **data is collected from diverse resources**, so to provide storage space great enough, reliable, and safe is a complex and structured process. New technologies like cloud computing services reveal a shift to a new computing paradigm, and it has become increasingly challenging to assure consistency in managing such large-scale data in cloud storage [53].

Local storage space of biobanks is or will be in a short time full, and the cloud storage system is becoming more and more important. With this, the problem of security and safety of data in clouds is rising, as well as the financial aspects and sustainability.

Establish a data preparation routine

1. Anonymization and resisty
2. Data harmonization
3. Metadata extraction
4. Cataloguing

Data Preparation Pipelines

- Data format Analysis
- De-identification
- Data Structure Harmonization
- Data Cleaning / Elimination of Unnecessary Data
- Metadata Extraction
- Data Transfer to Repository

- Metadata Publication

Data preparation routine

1. Anonymization and registry
2. Data harmonisation
3. Metadata extraction
4. cataloguing

Prepare appropriate infrasturcture

- Long term storage requirements
- Online access requirement
- Data preparation workstation
- Data preparation servers

Make Policies

- Ethical issues
- Data ownership
- Partnership
- Data availability

Ethical team

- Data preparation team
- Data specific advisor
- Disease specific advisor
- Engineering staff
- Biobank consultants

Adopt data structure standards

- Imaging data
- Biosignal data
- Tabular data
- Multimodal data

Document

- Medical data use agreements
- Partner center agreement
- Medical data donation agreement
- Informed consent for clinical data donation

Quality data

- Establish collaboration with clinical centres
- Promote data donation
- Conduct cohort studies

Tools

- Modality group specific data tepostitories
- Data preparation pipelines
- Data labelling tools
- portal

OPEN SOURCE IMAGE ANNOTATION TOOLS

1. 3D Slicer (www.slicer.org)
2. ePAD (epad.stanford.edu)
3. Horos Viewer (www.horosproject.org)
4. ImageJ/FIJI (fiji.sc)

5. InVesalius (invesalius.github.io)
6. ITK-SNAP (www.itksnap.org)
7. MedSeg (www.medses.ai)
8. MeVisLab (www.mevislab.de)
9. MITK 3D Slicer (www.mitk.org)
10. ParaView (www.slicer.org)
11. Seg3D (www.sci.utah.edu/cibc-software/seg3d)