

# Biobank metaportal to enhance collaborative research: [sail.simbioms.org](http://sail.simbioms.org)

## BIOBANKS

In order to identify new ways to prevent, diagnose and treat diseases, biobanks have been systematically collecting samples of human tissues and population-wide data on health and life-style. Such collections contain millions of tubes with biomaterial in a storage (freezer) and records about millions of people and thousands of measurements, often carried out in a longitudinal fashion. The outcomes of biobank-based studies are of great value for healthcare, academia and biomedical industry [McCarthy, M.I. *et al* , 2008, Yuille, M. *et al*, 2007]

## ACCESS TO BIOBANKS

Efficient access to population biobank data and to biomaterial is crucial for development and marketing of new pharmaceutical products, especially in the area of personalised medicine [Kauffman and Cambon-Thomsen, 2008]. However, due to ethical and legal reasons, biobanks are not at liberty to release the data or share biomaterial without approval of the local access committees. This leads to a “Catch22” situation when a biobank is not in a position to release any data until the purpose and design of the study is presented and approval is granted, while a party interested in the data is not able to apply for such access since it is not clear whether data in this particular biobank is suitable for the scope of a proposed study [Helgesson G. *et al* , 2007].

## SEMANTIC DIVERSITY OF BIOBANKS

Original sample annotation, captured at the time of collection, comes in a variety of formats and languages and there is no universal standard [Founti, P. *et al*, 2009, Fortier, I. *et al*, 2010]. Furthermore, various types of specialists (medical doctors, statisticians, geneticists and others) are accustomed to specific language (standard) [Hirtzlin, I. *et al*, 2003]. All this results in costly and repetitive data management: selective tagging, mapping and interlinking of various types of sample descriptions, i.e. *harmonisation*. Besides, descriptions (ontologies, controlled vocabularies, free text, semantic web concepts) come in a multitude of formats (RDF, XML, OWL).

## SOLUTION

We offer an informatics solution, that can assist in building efficient research collaborations through a controlled data release by biobanks to potential and existing partners.

SAIL ([sail.simbioms.org](http://sail.simbioms.org)), the Sample avAILability System, is an on-line resource, which allows researchers to locate and estimate the amount of relevant biomaterial available from a sample collection. SAIL provides information for each sample on whether a value for a given phenotypic variable exists or not,

without storing or disclosing the value per se. Phenotypic variables are organised in controlled vocabularies, taxonomic structures and studies.

The resource has been successfully used for retrospective harmonisation of phenotypic information from hospitals and biobanks, and it currently contains around 200 000 samples from 14 collections [Gostev M *et al*, 2011].

The SAIL mission as an online resource is to increase the visibility of the biobank content and to ease the set up of population-wide genetic and molecular studies.

#### References:

Fortier, I. *et al* (2010) Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *Int. J. Epidemiol.*, dyq139, 2010

Founti, P. *et al* (2009) Biobanks and the importance of detailed phenotyping: a case study-the European Glaucoma Society GlaucoGENE project. *Br J Ophthalmol*,93, 577-581.

Gostev M *et al*. (2011) [SAIL—a software system for sample and phenotype availability across biobanks and cohorts](#). *Bioinformatics.*, Feb15;27(4):589-91.

Helgesson G. *et al* (2007). Ethical framework for previously collected biobank samples. *Nature Biotechnology*, 25, 973-976.

Hirtzlin, I. *et al* (2003) An empirical survey on biobanking of human genetic material and data in six EU countries. *European Journal of Human Genetics*,11, 475-488.

Kauffman and Cambon-Thomsen (2008) Tracing Biological Collections: Between Books and Clinical Trials. *JAMA* 299, 2316-2318.

McCarthy, M.I. *et al* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Reviews Genetics*, 9, 356-369.

Yuille, M. *et al* (2007) Biobanking for Europe. *Brief Bioinform*, 9 (1), 14-24.