ESBB Working Group on Metadata: "ESBBperanto"

Background

Modern research projects require large numbers of biomaterials and associated data to produce meaningful results. Most biobanks cannot individually provide the required numbers leaving researchers to approach multiple biobanks to address their needs.

Biobank registries are useful to identify sample availability. MIABIS (Minimum Information About Biobank data Sharing) provides a mechanism for standardizing biobank catalogues. However, query results are often a heterogeneous mix of information and definitions. This is compounded in Europe where cultural and linguistic differences affect data.

What is required for an optimized exchange of biomaterials is standardized metadata. Metadata are literally 'data about data'. They provide information on such aspects as the 'who, what, where and when' of data and can be considered from the perspective of both the data producer and the data consumer.

For the producer, metadata are used to document data in order to inform prospective users of their characteristics, while for the consumer, metadata are used to both discover data and assess their appropriateness for particular needs – their so-called 'fitness for purpose'.

The problem of heterogeneity of information is two-fold. It needs to be addressed for retrospective and prospective data in all biobanking disciplines.

On the one hand, for prospective collections, a minimum data set must be defined. Definitions and nomenclatures are needed to describe data items. E.g.,within the field of human disease biobanking, Biospecimen Reporting for Improved Study Quality (BRISQ)[1] has provided a tiered list of items and a first step could be assigning nomenclatures or controlled vocabulary to the first tier containing essential data.

On the other hand, eetrospective collections require a mapping system to the defined minimum data set. A system similar to the Unified Medical Language System (UMLS) [2], National Institutes of Health, might be considered.

In the field of environmental biobanking, the Global Biodiversity Information Facility (GBIF) [3] promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet.

An important goal for GBIF is to develop the infrastructure needed across its network to support the management and delivery of the highest quality metadata that will enable potential end users to

[1] Moore et al. Cancer Cythopathology 2011;119:92-101
[2] http://www.nlm.nih.gov/research/umls
[3] www.gbif.org

easily discover which datasets are available, and, critically, to evaluate the appropriateness of such datasets for particular purposes.

Purpose

The biobanking community would benefit from a dedicated working group with experts from different biobanking fields, dedicated to harmonizing definitions and languages for biobanking.

The purpose of the ESBBeranto Working Group is to bring together ESBB members of various disciplines in order to identify, elaborate and launch common research projects.

This ESBBperanto working group will enable comparability of biomaterials and data across different biobanks and countries.

Goals

Biobanks and researchers would benefit from standardized information on biomaterials.

General concepts should be established and adapted to various biobanking fields, biomedical and environmental.

The goals of the ESBBperanto working group are to provide:

- a forum that brings together the various biobanking groups in a "translational" way regarding the concept of metadata;
- a working group dedicated to harmonizing procedures for biobanking data exchange;
- analysis of the state of data sets in place in the various disciplines;
- reports on potential options for data-sets such as BRISQ, GBIF, etc.;
- recommendations on definitions and nomenclatures for minimum data-sets;
- white papers on relevant biobank metadata including recommendations for ISBER Best Practices