

PSOM Biospecimen Annotation Standard for Medical Research

Under the leadership of Dr. Katherine Nathanson and Dr. Michael Feldman, a PSOM biospecimen annotation standard was developed and released in March 2016. The goal of the standard is to promote consistency across the institution in terms of specimen annotation and sharing of biospecimen data across medical research projects at Penn and with other institutions.

Biospecimen Annotation Standard

The standard was developed in a series of workshops attended by several in-house biospecimen lab managers, bio-informaticians, Dr. Christian Stoeckert (Genetics, IBI), and Drs. Nathanson and Feldman. It borrows heavily from the NCI CATissue package, the NCI CDEBrowser data, and established data annotations used by pre-existing repositories at Penn. All permissible values (for example, body site locations) are also mapped to ontologies to allow for flexible data analytics.

Although the first implementation is influenced the need for tissue annotation in the databases containing information about cancer specimens, an active stewarding process will allow the standard to be expanded to include additional annotations that other clinical groups may require, and to keep up with the changes in science and medicine.

Implementation

PSOM Biospecimen Review Committee

A PSOM Biospecimen Committee will review and approve changes to the standard. The PSOM Biospecimen Annotation Data Steward will be responsible for convening the committee on an ad hoc basis as decisions are required. The PSOM Biospecimen Review Committee will be appointed by the PSOM IT Council.

PSOM Biospecimen Annotation Data Steward

The process for the PSOM Biospecimen Annotation Data Steward will be put in place to provide a consistent method of managing necessary changes to the standard. The steward is the go-to resource when an update to the standards is proposed:

Changes in annotations:

- The LIMS team and/or PIs may request new annotations to be added to the standard. Only annotations which represent new concepts or provide for more detailed concepts will be considered. For example, if an annotation is required for time-out-of-refrigeration for a specimen and none exist, then the annotation can be proposed.

Changes in permissible values:

- The standard defines permissible values for many of the annotations. For example, permissible values are given for body sites and specimen processing method. Each permissible value is also mapped to an ontology term to allow for more flexibility in analyzing the data and to enable integration with data from other sources. For example, the specimen type "Plasma" is mapped

to the class “blood plasma specimen” in the Ontology for Biomedical Investigations. The steward will review with the LIMS team proposed additions to the permissible values and map the terms to the appropriate ontology.

Provisional Changes:

- Proposed annotations or permissible values can be provisionally approved by the data steward and implemented in the LIMS systems, and presented for formal approval to the PSOM Review Committee at its next meeting.

LIMS application

In 2012, PSOM invested in a commercial LIMS package from LabVantage, and is subsidizing its deployment and operation by PMACS. LIMS software provides a system that helps track biospecimens and biospecimen data through their full lifecycle, including sample collection, aliquot creation, derivative creation, sample movement, sample analysis, sample depletion, etc. The LIMS allows a centrally managed system to serve diverse research needs, and ensures research data is kept secure both internal and external to Penn. Going forward, the PSOM sponsored the PMACS LIMS team will configure all new biospecimen databases following the standard. Biospecimen databases that do not conform to the standard will not be supported through PSOM and PMACS unless granted an exception by PSOM Biospecimen Review Committee.

Retiring the caTissue and other legacy biospecimen databases at PSOM

Over 35 biospecimen collections are being tracked in the PSOM hosted caTissueⁱ database. At the request of the PSOM Senior IT Council, a PMACS project is being launched to migrate all the collections to LabVantage and as part of the process to adopt the new PSOM annotation standards. A critical success factor for the project is the participation of each affected laboratory to assist the PMACS team in that effort, in particular with the mapping of annotations and permissible values from caTissue to the PSOM Biospecimen Annotation Standard.

There are other laboratories at PSOM with biospecimen databases that may also choose to migrate to the PSOM LabVantage package; those laboratories will have to be ready to invest time in supporting the migration effort.

Abbreviations

PMACS - Penn Medicine Academic Computing Services

PSOM - Pennsylvania School of Medicine

LIMS - Laboratory Information Management Systems

ⁱ caTissue is the caBIG[®] tissue bank repository tool for biospecimen inventory, tracking, and basic annotation.