The NIH BD2K Center for Expanded Data Annotation and Retrieval

CEDAR

Most researchers understand the value of standardized descriptions when using third-party datasets; but when asked to structure and enrich the description of their own datasets, prior submission to databases, they view requests for even "minimal" information as burdensome.

CEDAR is studying the creation of comprehensive and expressive descriptions of datasets, using community standards, and lowering the bar for authoring metadata, in order to facilitate data discovery, interpretation, and reuse.

Our Approach

1 **Map** the landscape of content standards via biosharing.org

2 **Structure**

Develop methods for creating templates.

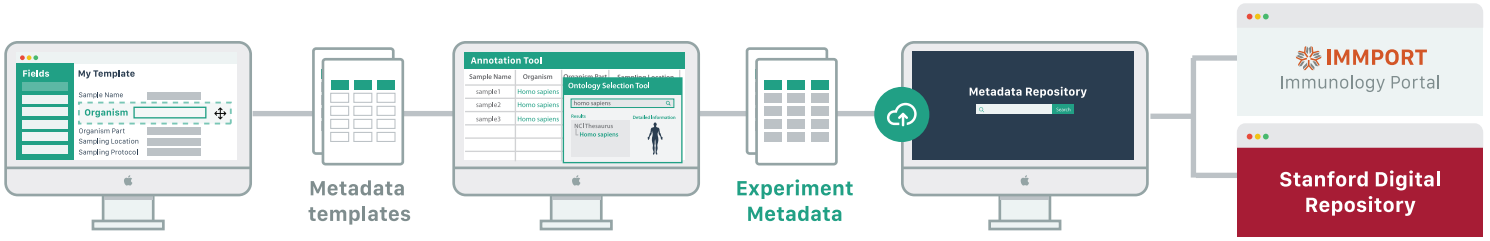
3 **Annotate**

Develop methods to ease the use of templates.

4 **Explore**

Create a repository of populated templates.

5 **Case Studies**



Authoring of metadata templates.

- Create and use 'elements' from content standards biosharing.org
- Create a language to represent relations among 'elements' W3C HCLS WGs
- Use existing examples of templates isatools IMMPORT HIPC

Annotation of data with metadata.

Exploration and reuse of datasets through metadata.

Facilitate submission of datasets to our two case study repositories and progressively to others.

Fill in template

Contribute

Submit, Search, & Reuse

Scientists

6

- Analyze the CEDAR repository of populated templates to reveal patterns in the metadata that will enable the metadata tools to use predictive data entry to ease the task of filling out the templates.
- Augment those metadata with links to the published literature (including secondary analyses and retractions).
- Augment those metadata with links to follow-up experiments (in online databases and in the literature).
- Allow the scientific community to comment on the experiment through structured metadata.

metadatacenter.org

A Center of Excellence supported by

NIH Big Data to Knowledge (BD2K)

under grant U54 AI117925 from the The National Institute of Allergy and Infectious Diseases

Stanford Medicine

Mark A. Musen (PI)
Carol A. Bean
Michel Dumontier
Olivier Gevaert
Purvesh Khatri
Martin J. O'Connor

Stanford University Libraries

Kim A. Durante

University of Oxford

Alejandra Gonzalez-Beltran
Philippe Rocca-Serra
Susanna-Assunta Sansone

Yale School of Medicine

Kei-Hoi Cheung
Steven H. Kleinstein

Northrop Grumman

Jeffrey A. Wiser