Supporting Ontology-Based Standardization of Biomedical Metadata in the CEDAR Workbench

Marcos Martínez-Romero, Martin J. O’Connor, Michael Dorf, Jennifer Vendetti, Debra Willrett, Attila L. Egyedi, John Graybeal, and Mark A. Musen
Stanford Center for Biomedical Informatics Research
Stanford University, Stanford, CA 94305, U.S.A.

CEDAR is developing novel methods and tools to create ontology-based metadata in biomedicine.

High-quality metadata are crucial for facilitating the discovery of scientific datasets and the reproducibility of the corresponding experiments. Despite the wide availability of biomedical ontologies, metadata submitted to public repositories rarely use standard terms. Consequently, finding scientific datasets and understanding the corresponding experiments can be extremely hard and time-consuming, and often requires post-processing.

The Center for Expanded Data Annotation and Retrieval (CEDAR) is creating a computational ecosystem for development, evaluation, use, and refinement of biomedical metadata. Our approach centers on the use of metadata templates (or templates), which define data elements needed to describe biomedical experiments.

The CEDAR Workbench provides three components that form a metadata construction pipeline: 1) a Template Designer, which supports template creation; 2) a Metadata Editor, to fill in templates with metadata; 3) a Metadata Repository for storing templates and metadata.

We extended the Template Designer and Metadata Editor to let users specify semantic content for templates and to easily enter ontology terms in their metadata.

Class and Property search: The Template Designer allows template authors to search for ontology classes (e.g. Publication) and properties (e.g. hasTitle) to annotate their templates.

Value Set creation: Template authors can define lists of ontology terms to constrain field values (e.g. Prospective study, Retrospective study, and Hybrid study for a study type).

Class creation: CEDAR allows users to define new classes and immediately use them.

Class and Property Search: The Template Designer allows template authors to search for ontology classes (e.g. Publication) and properties (e.g. hasTitle) to annotate their templates.

Value Set creation: Template authors can define lists of ontology terms to constrain field values (e.g. Prospective study, Retrospective study, and Hybrid study for a study type).

Class creation: CEDAR allows users to define new classes and immediately use them.

Screen shot of Template Designer’s ontology lookup interface. The user searched for sleep disorder and selected the term from DOID.

Screen shot of the Metadata Editor that shows the possible values of a Disorder field in a Study template. This field has been constrained to accept values from the branches of the DOID ontology with roots cognitive disorder, sleep disorder, and dissociative disorder.

- Increases metadata quality
- Reduces authoring effort
- Generates reusable metadata

Screen shot of the Metadata Editor that shows the possible values of a Disorder field in a Study template. This field has been constrained to accept values from the branches of the DOID ontology with roots cognitive disorder, sleep disorder, and dissociative disorder.

Value Constraints: Template authors can constrain field values to (1) ontologies; (2) ontology branches; (3) specific classes; and (4) value sets.

The CEDAR Workbench offers a simple approach to generate ontology-based metadata.

CEDAR Workbench: cedar.metadatacenter.net

metadatacenter.org

CEDAR is supported by grant U54 AI117925 awarded by the National Institute of Allergy and Infectious Diseases through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative.