FAIR LINCS Metadata Powered by CEDAR Cloud-Based Templates and Services

Daniel Cooper[1,2,3,5], Amar Koleti[1,2,5], Martin O'Connor[4,5,6], Debra Willrett[4,5,6], Caty Chung[1,2,5], John Greybeal[4,5,6], Mark Musen[4,5,6], Stephan Schürer[1,2,5]

[1]Center for Computational Science, University of Miami, Miami, FL, USA; [2]Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL, USA; [3]Miami Project to Cure Paralysis, University of Miami Miller School of Medicine, Miami FL, USA; [4]Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA; [5]BD2K LINCS Data Coordination and Integration Center; [6]BD2K CEDAR The Center for Expanded Data Annotation and Retrieval

Abstract

The Library of Integrated Network-based Signatures (LINCS) program generates a wide variety of cell-based perturbation-response signatures using diverse assay technologies. For example, LINCS includes large-scale transcriptional profiling of genetic and small molecule perturbations, and various proteomics and imaging datasets. We currently obtain metadata through an online platform, the metadata submission tool (MST), based off the use of spreadsheet data templates. While functional, it remains difficult to maintain FAIR standards, specifically remaining findable and re-usable, for metadata without (enforced) controlled vocabulary and internally built linkages to ontologies and metadata standards. To maintain FAIR-centric metadata, we have worked with the Center for Enhanced Data Annotation and Retrieval (CEDAR), to develop modular metadata templates linked to ontologies and standards present in the NCBO Bioportal. We have also developed a new LINCS Dataset Submission Tool (DST), which links new LINCS datasets to the form-fillable CEDAR templates. This metadata management framework supports authoring, curation, validation, management, and sharing of LINCS metadata, while building upon the existing LINCS metadata standards and data-release workflows. Additionally, the CEDAR technology facilitates metadata validation and testing, enabling users to ensure their input metadata are LINCS compliant prior to submission for public release. CEDAR templates have been developed for reagent metadata, experimental metadata, to describe assays, and to capture global dataset attributes. Integrating the submission of all these components into one submission tool and workflow we aim to significantly simplify and streamline the workflow of LINCS dataset submission, processing, validation, registration, and publication. As other projects apply the same approach, many more datasets will become cross-searchable and can be linked optimizing the metadata pathway from submission to discovery.

Materials and Methods

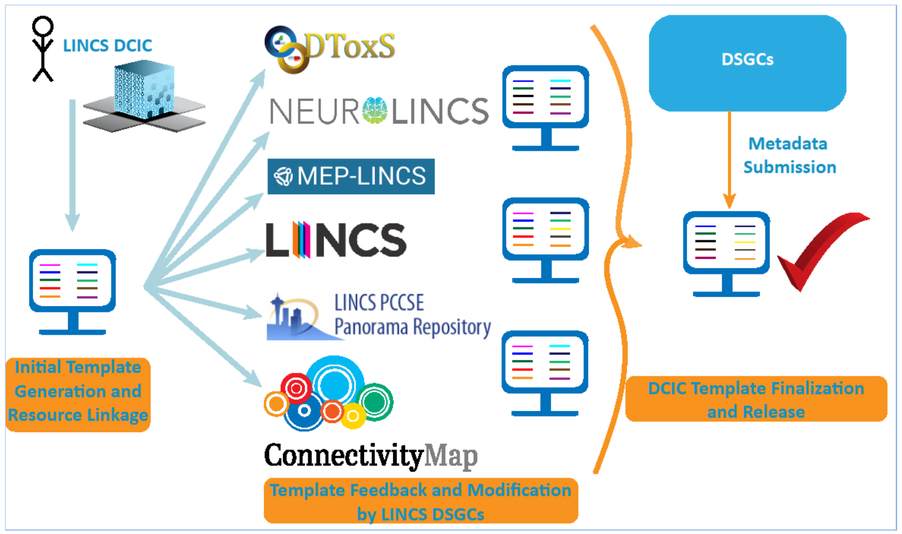

Figure 1 – Building and establishment of metadata standards within the LINCS consortium. Initial standards are proposed to consortia members, then feedback is consolidated and implemented for use by members for metadata submission

Results

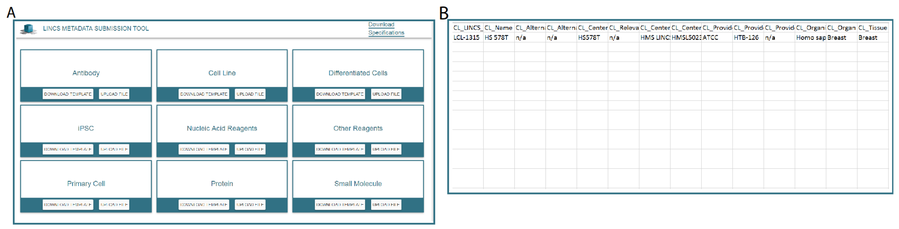

Figure 2 – LINCS Metadata Submission tool. Current paradigm for submission of metadata to LINCS system is through the MST (A), based on simplified spreadsheet templates (B).

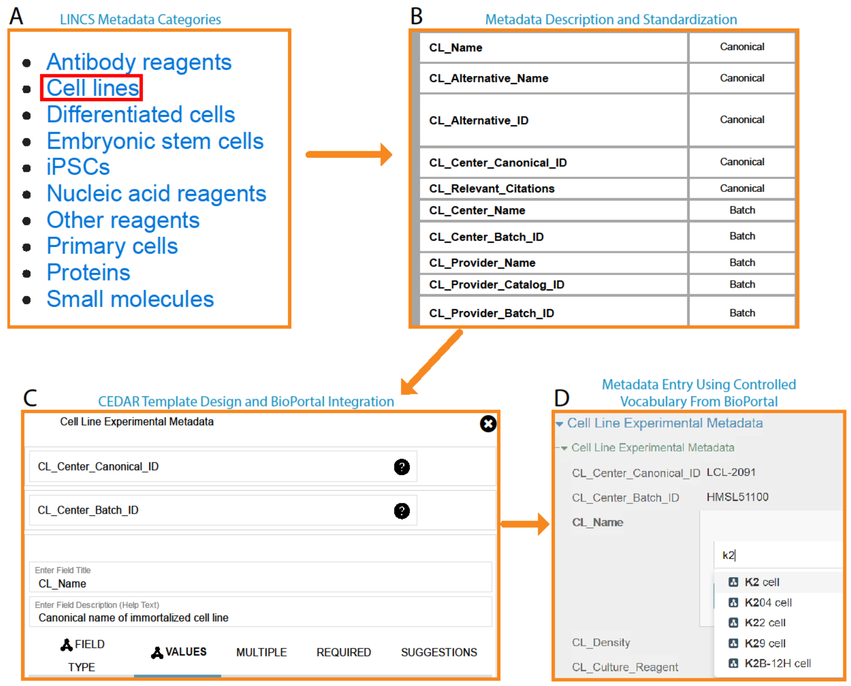

Figure 3 – Building and utilization of LINCS metadata templates in the CEDAR framework. LINCS metadata standards (A and B) are input to the CEDAR system as templates linked to BioPortal ontology values (C) to establish and maintain controlled vocabulary within field names and content (D).

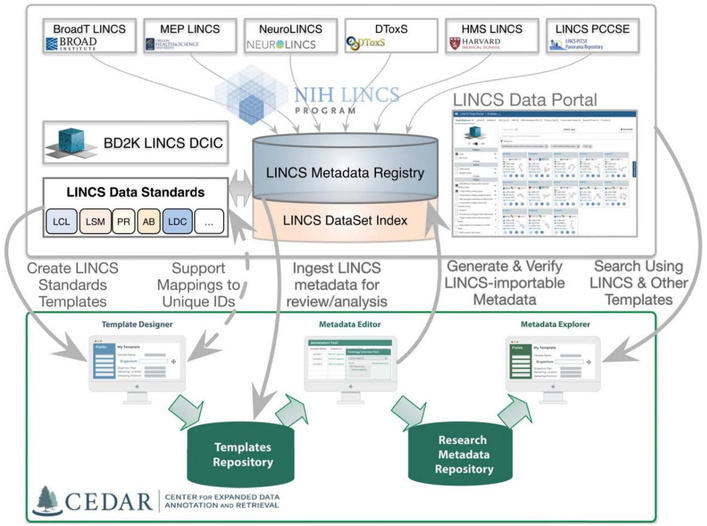

Figure 4 – Integration of LINCS data repository and CEDAR metadata framework. LINCS metadata standards are input to CEDAR framework and completed by data generation centers. Datasets are submitted directly to the DCIC, where they are packaged and associated with appropriate metadata entity batches.

Figure 5 – LINCS Data Submission Tool. Enables submission of dataset, assay, and entity level data through direct integration with CEDAR.

Conclusions

- **New metadata standards established by LINCS consortium**
- **Key to FAIRness utilizes CEDAR linkages to external persistent IDs (PURLs)**
- **CEDAR framework enables simplification and streamlining of metadata submission workflow and FAIR maintenance**

Acknowledgements: This research was supported by grant U54HL127624 awarded by the National Heart, Lung, and Blood Institute through funds provided by the trans-NIH Library of Integrated Network-based Cellular Signatures (LINCS) Program (http://www.lincsproject.org/) and the trans-NIH Big Data to Knowledge (BD2K) initiative (http://www.bd2k.nih.gov).

Correspondence: s.schurer@miami.edu

BD2K-LINCS
DATA COORDINATION AND INTEGRATION CENTER