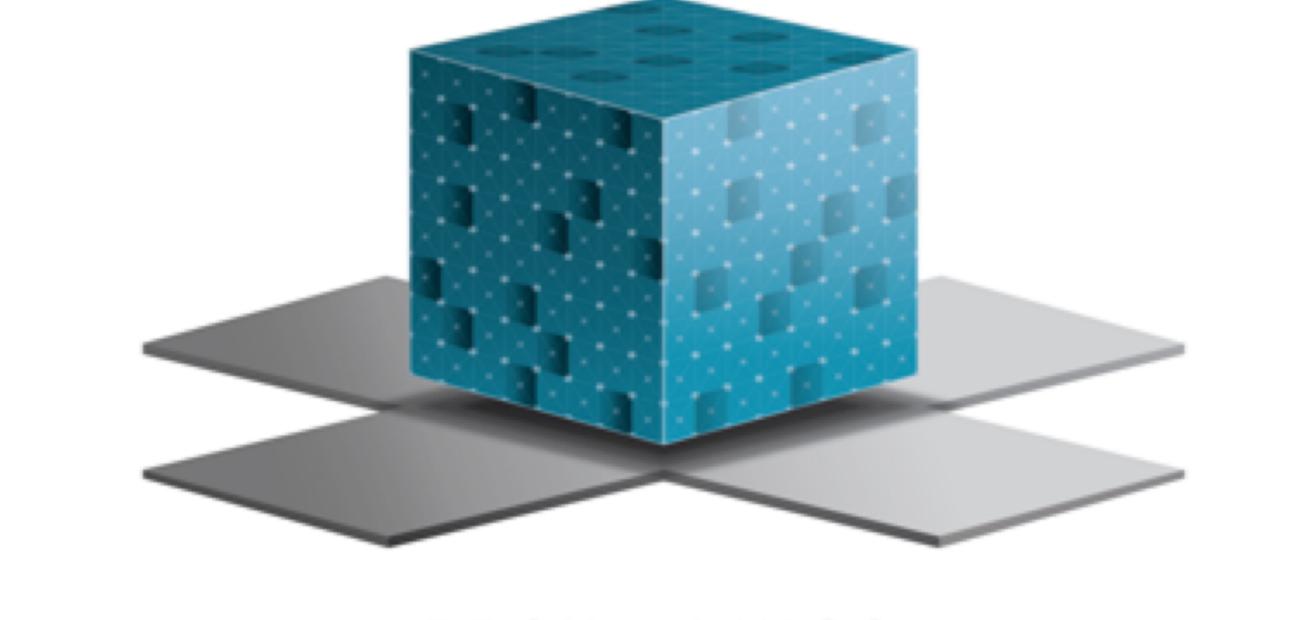
## LINCS Dataset Registry (LDR): A Web-Based System to Capture and Manage LINCS Data Releases with Autocomplete Web Forms

Available at http://amp.pharm.mssm.edu/LDR

Michael G. McDermott, BS<sup>1,2</sup>, Qiaonan Duan<sup>1,2</sup>, Amar Koleti<sup>2,3</sup>, Dusica Vidovic<sup>2,3</sup>, Stephan Schurer, PhD<sup>2,3</sup>, Chris Mader<sup>2,3</sup>, Michael J. McCoss<sup>4</sup>, Avi Ma'ayan, PhD<sup>1,2</sup>





DATA COORDINATION AND INTEGRATION CENTER

<sup>1</sup>Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1215, New York, NY 10029 USA; <sup>2</sup>BD2K-LINCS Data Coordination and Integration Center (DCIC); <sup>3</sup>Center for Computational Science, University of Miami, Miami, FL, USA; <sup>4</sup>Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA

## Abstract

One of the challenges of BD2K is to capture metadata about dataset instances and link such metadata to controlled dictionaries and ontologies. This metadata capture is expected to improve dataset search and facilitate data integration. This challenge is central to the LINCS program and the BD2K-LINCS DCIC because different LINCS data generation centers use different but overlapping assays, perturbations (genes/proteins, small-molecules), cell-lines, disease models, readouts and other common entities. Most major data repositories such as GEO or Chorus currently do not have advanced webbased forms to capture metadata about dataset instances from their data submitters. In year 1, the BD2K-LINCS DCIC developed the LINCS Dataset Registry (LDR) system to capture, visualize and manage all LINCS released datasets. LDR is a modern, mobile-friendly web application designed to streamline the process of submitting, approving, and releasing datasets. LDR consists of a client-side application created with the JavaScript library AngularJS and a web server application written in NodeJS. The server's extensive API's communicate to a MongoDB database responsible for storing and querying each center's data. LDR has login authentication functionality that enables the security of unreleased datasets, and its advanced input forms allow for fast, hassle-free data entry. Form entities have autocomplete functionality drawing from ontologies and dictionaries managed by live remote servers. LDR also contains a dataset-specific message board that enables communication between the LINCS data generation centers and the NIH staff to facilitate an approval process. While designed for LINCS, LDR will be generalized to facilitate data capture for other projects. For a BD2K Common Fund supplement in collaboration with the McCoss Lab at the University of Washington, we will reuse the LDR code as an additional feature for Chorus, a new cloud-based application that provides scientists with the ability to securely store, analyze and share their MS data regardless of the original raw file format. The goal of Chorus is to create a complete catalogue of the world's mass spectrometric data that can be openly accessed by, and freely accessible to, the global scientific community as well as the general public. Hence, the web-forms, developed for LINCS, will be provided as a webservice and will be embedded within Chorus. This effort will serve as a model for the Chorus/proteomics community to better annotate their proteomics data submissions but will also introduce a global solution for other repositories.

## Methods

The LINCS Dataset Registry is composed of a web application written in HTML, CSS, and Javascript (AngularJS), a web server written in NodeJS that interacts with the metadata registry hosted at the University of Miami in addition to its own MongoDB database. The interaction of these components is shown in figure 1.

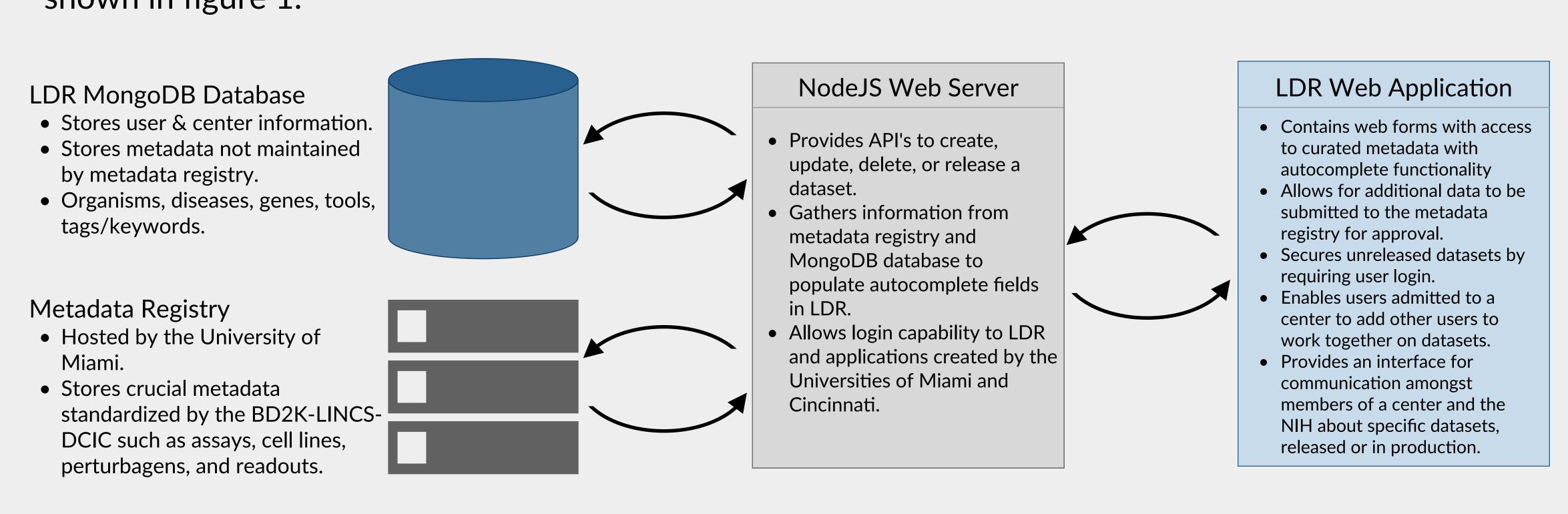


Figure 1. The LINCS Dataset Registry's architecture: A front-end web application, NodeJS web server, MongoDB database, and external metdata registry.

Results

In order to use the Harmonizome mobile application, a device running Android (4.1+) or iOS (8.2+) with an internet connection is required. The application is downloaded from the Google Play Store for Android devices or the App Store for iOS devices. Once the application is started, the screen will look similar to figure 2. By typing a query into the search bar, you'll notice that the appearance will change to that of figure 3. The terms available in the autocomplete options below the search bar are pulled from the Harmonizome's web server. While you are required to select one of these provided terms, there are over 28,000 available and new ones are being added. Finally, selecting a term brings you to the category screen, where results are sorted by various categories depending on the term selected. This view is shown in figure 4.

Conclusions

References

Acknowledgements

This work is supported by NIH grants: R01GM098316, U54HG008230 and U54CA189201 to AM.