Clean-Slate Development of a Hydra Repository Service Infrastructure

John Bell < john.p.bell@dartmouth.edu> | Carla Galarza < carla.m.galarza@dartmouth.edu>



The Plar

Connected Applications PLUM LNA ACADEMIC COMMONS Digital Repository Harvests citations from Linked Name Authority: Long-term storage and Generates advanced usage Secure storage and ORCiD: Unique identifier external sources, stores full discovery of OA-compliant management of the Library's metrics from social media public data source for lookup and citation syndication digital collections. Dartmouth-specific entities documents and objects sources and links bibliographic record Data sources, HR feed, HR feed, student feed, LNA feed, librarian Current CONTENTdm assets, Elements citation feed, support, file uploads, LNA citation feed ORCiD citation lists workflow for citation metadata unique IDs metrics data new digitization projects Aggregated name lists, File and metadata Bibliographic record for Digital assets' metadata, Faculty citation lists, citation information, and Citation and usage metrics unique identifiers derivatives and master files supporting OA policy traditional outputs unique IDs Machine-readable identifiers Public discovery and Flexible UI for patrons and Bibliographic reports, Unique identifiers, Data for metrics reporting and data for Dartmouth storage interface, data staff for all of the Library's curated citation lists and discovery tools citation lists feeds for stored objects people and citations digital collections

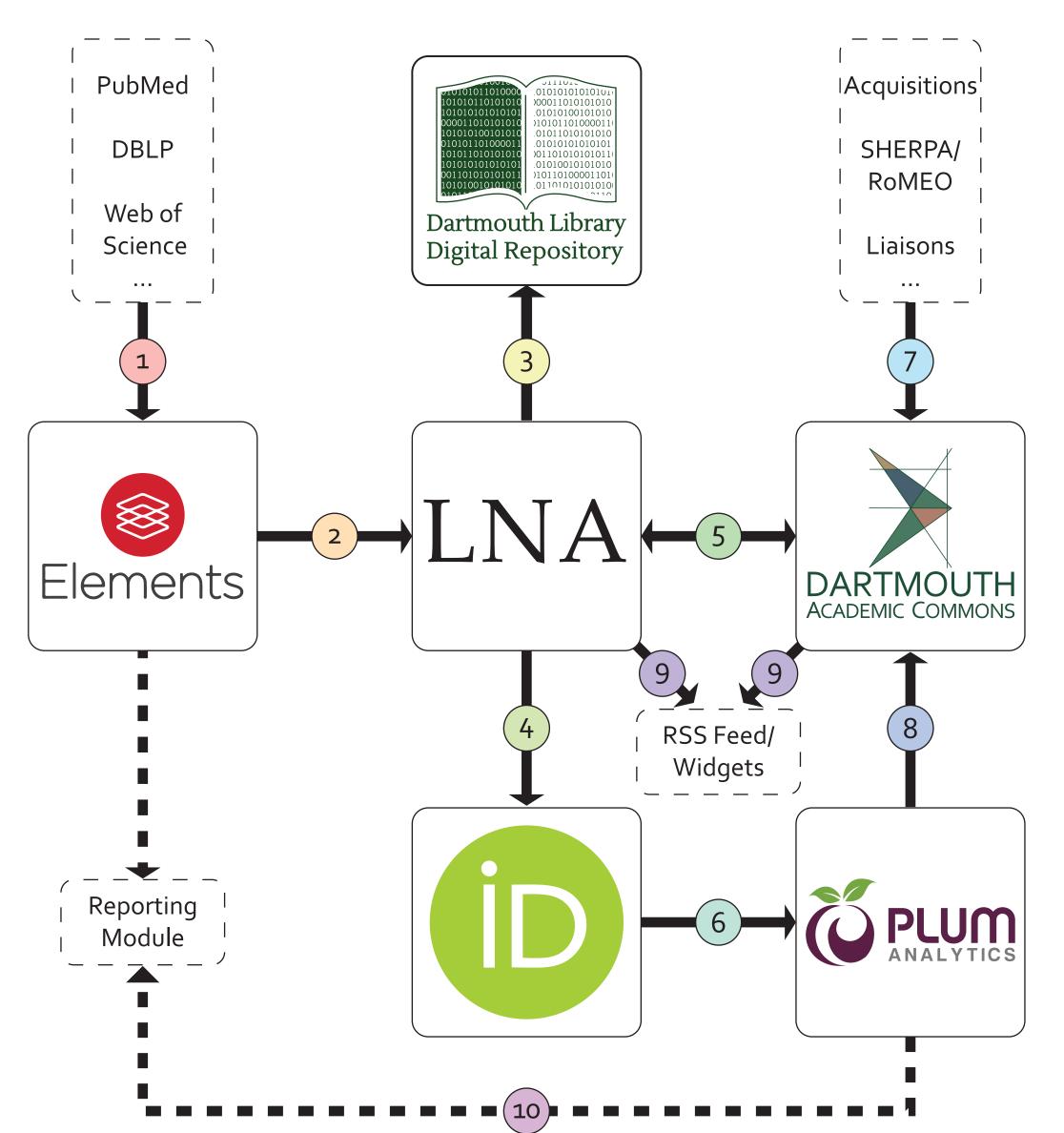
Goals

- ➤ Multiple Hydra/Fedora4 installs to fulfill distinct functional niches as standalone applications
- > Each application uses common identifiers managed in Linked Name Authority
- > Applications connected by service bus managing public and secure API calls
- ➤ Federated search across Open Access media (DAC), bibliographic record of Dartmouth-sourced articles in external stores (LNA), and public and private Library collections (DLDR)

Features

- ➤ Acquisition, ingest, and rights management workflow managed across related Hydra/Fedora systems using GTD-style interface
- > Librarian mediated submission of Open Access content
- Automatic syndication of faculty bibliographic record as RSS/Atom
- ➤ Algorithmic discovery and suggestion tools based on keyword, faculty research interests, internal usage metrics and altmetrics from a variety of sources that are determined by academic departments

Article Data Path

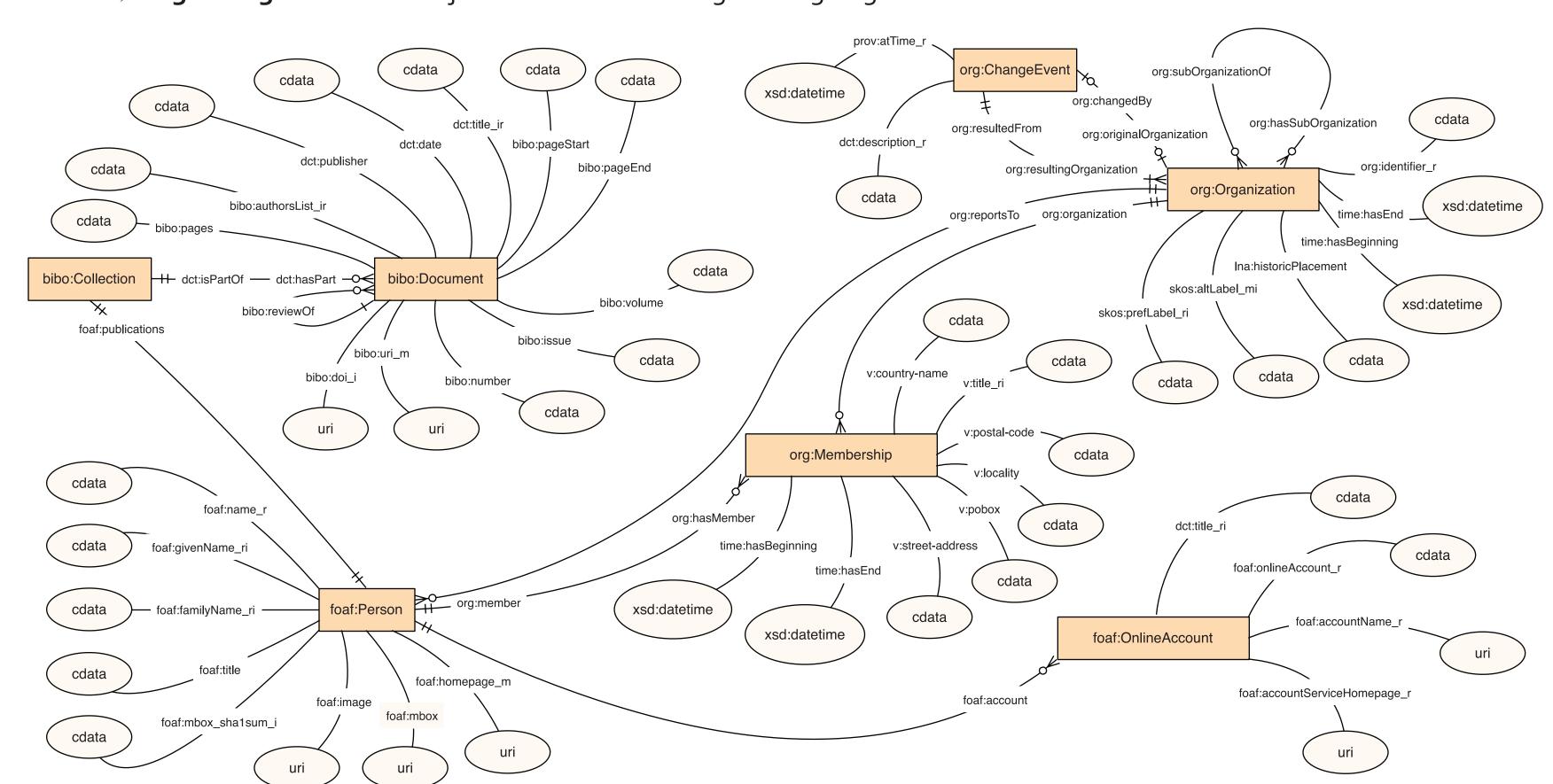


- Elements harvests some citations from external sources based on string searches. Result is a list of potential citations that requires human curation and completion to reflect complete citation list.
- LNA ingests curated list of citations from Elements. Initial scope includes articles, monographs, conferences.
- LNA combines Elements data feed with data from human resources, the student database, and historic sources for RDF entity publication. One consumer is the Digital Repository.
- LNA uploads list of verified publications to ORCiD, populating ORCiD faculty citation profile.
- DAC harvests OA compatible publication list from LNA and work is queued for ingest if possible. For works where the canonical host is DAC itself, LNA pulls citation from DAC and updates ORCiD.
- Plum Analytics automatically updates its search parameters when ORCiD is updated.
- Tibrary sources upload locally-produced works into DAC, pulling metadata from LNA (where the citation should already exist), SHERPA/RoMEO, and others
- DAC collects and transforms impact data from Plum, using it to enhance discovery.
- DAC and LNA publish data feeds for faculty and departments that can be integrated into profiles.
- Plum impact metrics and Elements citation records are combined in an external reporting module.

Step 1: LNA

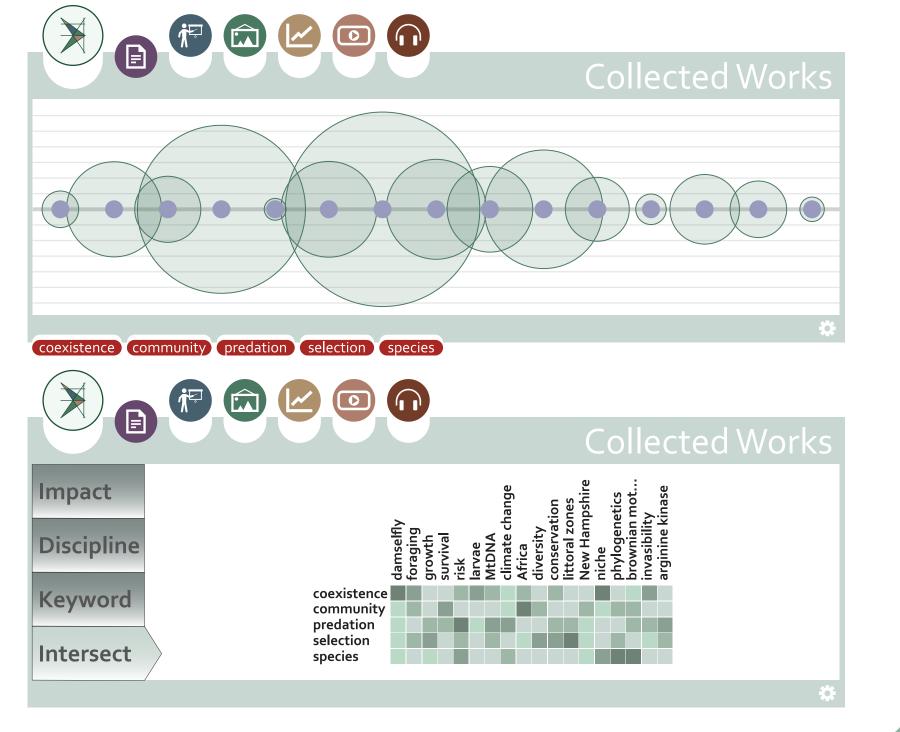
The Dartmouth Linked Name Authority (LNA) is an infrastructure project intended to support local and external systems that need Dartmouth-sourced entity metadata to operate. LNA is designed to fulfill and expand the role of the library-maintained authority file in a linked data environment. It will aggregate name and citation data from multiple siloed systems spread across Dartmouth's departments and enhance the available information as necessary to meet new, more stringent standards for data quality and completeness. The newly standardized data will then be released as RDF and searchable via an API upon which systems both internal and external to Dartmouth College will be able to draw, enhancing Dartmouth's ability to participate in the emerging linked data and federated information system ecology.

- > foaf:Person faculty, staff, or student associated with Dartmouth, either today or historically
- > bibo:Document an individual citation object, which may or may not be canonically located at Dartmouth
- > bibo:Collection a container for an individual's bibo:Document list
- > org:Membership an appointment or position held by a foaf:Person
- > foaf:OnlineAccount a unique online identifier associated with a foaf:Person (ORCiD, etc.)
- > org:Organization an organizational unit associated with Dartmouth, either today or historically
- > org:ChangeEvent an object used to track changes in org:Organizations over time



DARTMOUTH ACADEMIC COMMONS

After the initial release of the Linked Name Authority, development of the Dartmouth Academic Commons will begin. DAC is being designed around a visual search, facet, and discovery interface that contextualizes individual artifacts within a larger field of inquiry. Data will be passed to the DAC interface using JSON-LD and rendered graphically using D3.js. In addition to being displayed in search and browse interfaces, these visualization widgets will also be present in individual item displays. The goal is to reduce the repository's dependence upon discovery by external search engines and provide item-level contextual information for users who do arrive from a search engine.



Up Next