

UCLA

Publications

Title

IDRE Proposal: UCLA Data Registry System

Permalink

<https://escholarship.org/uc/item/824450x7>

Authors

Borgman, Christine L.

Grappone, Todd

Strong, Gary

et al.

Publication Date

2011-04-11

IDRE PROPOSAL: UCLA DATA REGISTRY SYSTEM

PI: Christine L. Borgman, Professor, Information Studies

Supporting Investigators:

- UCLA Library: Todd Grappone, Associate University Librarian, Susan Parker, Deputy Director, Gary Strong, University Librarian
- Center for Embedded Networked Sensing: Jeffrey Goldman, PhD, Administrative Director
- Department of Information Studies: Jillian Wallis, PhD Candidate

In 2009, the Center for Embedded Networked Sensing (CENS) began a pilot effort to register research data products, report their existence to responsible funding agencies, and make these data more publicly available. The data registry effort is working well on a small scale, and has the potential to serve the entire UCLA community if made more robust and feature rich.

The Data Registry Opportunity

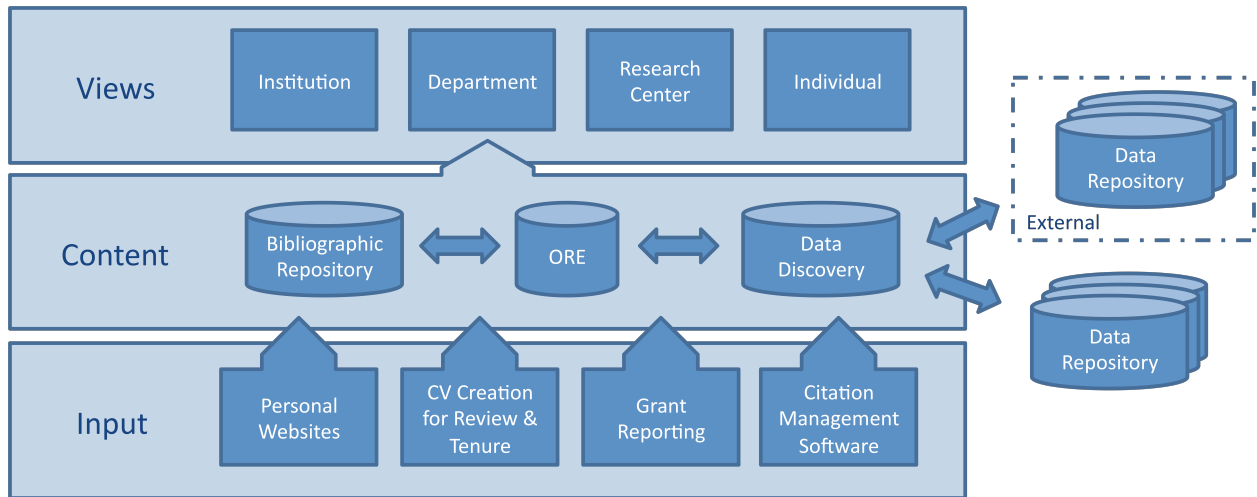
CENS, a very successful NSF Science and Technology Center, and UCLA's first NSF STC, officially closes on July 31, 2012. While we hope CENS will continue in some form, we have barely a year left to capture the legacy of data and publications from the Center. The UCLA Library is in a position not only to migrate the small data registry that CENS has created, but to use CENS's content and experience as the basis for a sustainable data registry. The registry will facilitate discovery of types of research data at UCLA and will support collaboration. The UCLA library, which is developing a Data Management Planning Tool, will incorporate the Data Registry into its larger plans for data repositories.

The UCLA Library has begun a project to migrate from their current Digital Asset Management (DAM) system to join an open source digital library project called Islandora. Islandora uniquely combines the [Drupal](#) and [Fedora](#) open software applications to create a robust digital asset management system that can be fitted to meet the short and long term collaborative requirements of digital data stewardship. The proposed project extends that effort to provide a repository and metadata management for the CENS data registry and to develop a campus-wide data registry.

Fedora (Flexible Extensible Digital Object Repository Architecture) was originally developed by researchers at Cornell University as an architecture for storing, managing, and accessing digital content in the form of digital objects. The software defines a set of abstractions for expressing digital objects, asserting relationships among digital objects, and linking "behaviors" (i.e., services) to digital objects and provides a core repository service (exposed as web-based services with well-defined APIs). In addition, Fedora provides an array of supporting services and applications including search, OAI-PMH, messaging, administrative clients, and more. Fedora provides RDF support and the repository software is integrated with semantic triple store technology, including the Mulgara RDF database. Fedora helps ensure that digital content is durable by providing features that support digital preservation. In addition to supporting international data exchange standards, this software has a proven track record and is well suited to meet the registry needs for much of campus and specifically the CENS registry data. The CENS registry, as a local pilot, does not support the technical standards necessary for interoperability and discovery.

Figure 1 outlines the components of the proposed data registry. The input layer is comprised of existing practices that could be harnessed to populate the content databases. The content layer has 3 main

databases and some ancillary databases that would be linked to through the use of Fedora-supplied RDF and ORE services. The Data Discovery database maintains a registry of datasets that have been created by members of the UCLA community. The datasets themselves are likely to be located elsewhere, either within the UCLA network or externally. A Bibliographic repository is a registry of publications, and in some cases provides the full text, when copyright allows, which in this case help to provide context, methods, and interpretations of the datasets in the Data Discovery registry. The Object Reuse & Exchange protocol (ORE) database explicitly links related papers and datasets. Data is served from the databases using views at different levels from the institution to the individual, allowing for dynamic reorganization of the materials depending on the user's interest.



The Islandora project undertaken by the Library will provide a development platform for the UCLA Data Registry. Furthermore the Fedora portion of the project will deploy web crawling technology (see: [Fedora Generic Search](#)) allowing for updating local bibliographic data by syncing and linking information with the published record.

Policy, system, and practices working together

This registry will allow datasets to be aggregated at different levels, such as all those datasets attributed to an individual, research center (such as CENS), department, or institution, to create a unified presence. This approach respects the varying practices by discipline and research specialty. In the longer term, the system could be extended to include statistics on data and publications by individuals and academic units and on frequency of access to these records. The UCLA Data Registry will be built upon open standards allowing for rapid integration to all existing and future web based administrative or research workflows on campus.

Workplan

Task	Month 1-3	Month 4-6	Month 7-9	Month 10-12
Analysis of current registry and survey of campus data efforts				
Formulation of data registry specifications				
Create a report summarizing the registration and access needs				
Migrate existing CENS data and develop registry UI				
Interface development with eScholarship				
Develop analytics interface				
Pilot testing with multiple campus groups, adjust and modify UI and backend as needed (GSR & PA III)				
Assessment of results				
Report and recommendations to campus				
		PA III		
		GSR		

Budget

The project will require a GSR Step VII for research and functional requirements development, under the direction of Professor Borgman. Additionally, a Programmer Analyst III, under the direction of the Library at 50% effort, will be needed for the project development. This proposal is asking for \$ 85,404 to cover the costs of the GSR and the PA III.

Budget and budget justification:

July 1, 2011 to June 31, 2012

Staff

GSR VII fee remissions, 3 terms	\$12,222
GSRVII information studies, 12 months, 49%	\$23,358
Programmer Analyst III, 12 months, 50%	\$49,824
Total	\$85,404

Funds are requested for one Graduate Student Researcher (step VII), at 49% for 12 months, starting on July 1, 2011. The Information Studies student will be an expert in metadata and in data. He or she will be reporting to Professor Borgman and will work with library staff to inventory data types most likely to be registered, develop the metadata structures, and map them to the appropriate technical standards. A fixed term PA III contract position to develop the data registry user interface from the specifications developed by the GSR, migrate existing registry data into Fedora, create an analytics dashboard for reporting on registry use and data and interface the data registry with eScholarship. The work will be under the direction of Library IT, will be supplemented as needed by existing library personnel and will reside on Library infrastructure.