Worlds Collide: A Repository Based on Technical and Archival Collaboration

Erin O'Meara
Electronic Records Archivist, University of North Carolina at Chapel Hill

Gregory Jansen
Repository Developer, University of North Carolina at Chapel Hill

The failure of many institutional repositories (IR) to acquire large sets of faculty publications has shown that the traditional IR model is not sustainable without a shift in academic publishing. The Carolina Digital Repository (CDR) aims to be more than a traditional IR and instead of focusing primarily on open access publishing, it will acquire, preserve and make accessible a range of at-risk scholarly output, such as datasets, faculty papers, university records and other faculty research projects.

Research libraries have begun to focus on their special collections as core to their mission and a key contribution to the wider world of scholarship. We extend this emphasis on collecting unique materials to include the new IR as a strategic direction.

Repository implementations suffer from a disconnect between technologists and digital curators in terminology and approach. This barrier has inhibited the development of comprehensive solutions that meet technical needs and deeply incorporate curatorial requirements. Overcoming this barrier demands work to establish common language and mutual understanding of concepts from each discipline.

The key output of academic institutions, besides the conferring of degrees, is the development of new knowledge. Through preservation a repository supplies a key missing ingredient in the life cycle of knowledge production, a persistent point of access that fills the growing gulf between modern modes of digital scholarship and traditional publishing. The new IR model addresses the institution's role in providing stewardship of these assets.

The Carolina Digital Repository (CDR) is an institution-wide initiative at University of North Carolina at Chapel Hill (UNC-CH), spearheaded by University Libraries. It is a preservation framework for material in electronic formats produced by members of the UNC-CH community. Following standards developed in the OAIS reference model, the CDR employs Fedora for data content models and uses iRODS as a data store.

In this paper we will explain how the repository works as a whole to meet these challenges. We will show how our repository was made possible through a sustained commitment of institutional resources and the deep involvement of motivated experts with archival, librarianship and software development backgrounds.

We will trace how the functional needs for preservation activities were brought into the project and satisfied without sacrificing flexible access and ingest work flow. We will explain how we blended and extended the models set forth in Open Archival Information System (OAIS), Fedora Commons and the Integrated Rule-Oriented Data System (iRODS) to meet these requirements without creating a muddle of concerns in code and subsystems. Lastly, we will explore the stable measures taken and the ongoing project to create data management and object integrity safeguards used to mitigate risk of data corruption and loss.

A key recommendation from the CLIR report for Reconceiving Research libraries for the 21st century is:

> Collaboration should under-gird all strategic developments fo the university, especially at the service function level. Greater collaboration among librarians, information technology specialists, and faculty on research project design and execution should be strongly supported. (CLIR 2008, 10)

The CDR has been a collaboration of four partners, University Libraries, Information Technology Services, the School of Information and Library Science and the Data Intensive Cyber Environments (DICE) research group. From conception to deployment, on committees, the project team and working groups, the initiative incorporated expertise from all of these disciplines.

The CDR's design is guided by archival theory that includes principles of authenticity and trust. OAIS provided a conceptual model that informed the design of system layers and components. Certification tools, such as Trustworthy Repository Audit Checklist (TRAC), supplied functional requirements. The resulting implementation is a distinct best-of-breed software architecture aimed at teasing apart the concerns of ingest work flow, durable access and data preservation. These worlds met as we adopted and adapted numerous preservation standards, including OAIS, METS, MODS and PREMIS.

The CDR is set to go live in April 2010. We will have several pilot collections populating the repository that span disciplines and collecting efforts. Our goals for the first year of deployment include an enhanced user interface with greater search and access control, enhanced metadata and curation of objects, and audit of current preservation activities within the repository.

References

Council on Library and Information Resources. No Brief Candle: Reconceiving Research Libraries for the 21st Century. August 2008.