Developing a Common Submission System for ETDs in the Texas Digital Library

Adam Mikeal, Tim Brace, John Leggett, Mark McFarland

Texas A&M University Libraries, The University of Texas Libraries

Abstract

The Texas Digital Library is a consortium of universities organized to provide a single digital infrastructure for the scholarly activities of Texas universities. The four current Association of Research Libraries (ARL) universities and their systems comprise more than 40 campuses, 375,000 students, 30,000 faculty, and 100,000 staff; while non-ARL institutions represent another sizable addition in both students and faculty[1]. TDL's principal collection is currently its federated collection of ETDs from three of the major institutions; The University of Texas, Texas A&M University, and Texas Tech University. Since the ARL institutions in Texas alone produce over 4,000 ETDs per year, the growth potential for a single state-wide repository is significant.

To facilitate the creation of this federated collection, the schools agreed upon a common metadata standard represented by a MODS XML schema[2]. Although this creates a baseline for metadata consistency, there exists ambiguity within the interpretation of the schema that creates usability and interoperability challenges. Name resolution issues are not addressed by the schema, and certain descriptive metadata elements need consistency in format and level of significance so that common repository functionality will operate intuitively across the collection.

It was determined that a common ingestion point for ETDs was needed to collect metadata in a consistent, authoritative manner. A working group was formed that consisted of representatives from five universities, and a state-wide survey of the state of ETDs was conducted, with varied levels of engagement with ETDs reported. Many issues were identified, including policy questions such as open access publishing, copyright considerations and the collection of release authorizations, the role of infrastructure development such as a Shibboleth federation for authentication, and interoperability with third-party publishers such as UMI. ETD workflows at six schools were analyzed, and a meta-workflow was identified with three stages: *ingest*, *verification*, and *publication*. It was decided that Shibboleth would be used for authentication and identity management within the application.

This paper reports on the results of the survey, and describes the system and submission workflow that was developed as a consequence. A functional prototype of the ingest stage has been built, and a full prototype with Shibboleth integration is slated for completion in May of 2007. Demonstrators of the application are expected to be deployed in fall of 2007 at three schools.

---

[1] John Leggett, Mark McFarland and Drew Racine. "The Texas Digital Library: A Business Case". Prepared and published by the Texas Digital Library, July 2005, revised July 2006.

[2] Surratt, Brian. "MODS Meets Manakin: Innovations in the Texas Digital Library's Thesis and Dissertation Collection". *Proceedings of the 9th International Symposium on Electronic Theses and Dissertations*. June 7—10, Quebec City, Canada