

NARRATIVE

Introduction

Theses and dissertations play a critical role in higher education. A necessary component for virtually all doctoral and many masters programs, these demonstrate the new scholar's ability to conduct independent, original, and meaningful research. The doctoral dissertation in particular is recognized for its significance in higher education. In *The Role and Nature of the Doctoral Dissertation: A Policy Statement*, the Council of Graduate Schools states "There is no question that, in the view of the faculty, students, and administrators participating in this study, the doctoral dissertation, as a demonstration of a student's ability to carry out research independently, defines the essence of the Ph.D. degree."¹

Traditionally, theses have been printed, bound, and stored in university libraries and archives. While the print model is well established, it also has many shortcomings. Print theses do not enjoy high circulation rates, and consequently work done by graduate students is not widely disseminated. Typically, only one or two bound copies are produced and stored locally, so holdings are limited to one institution. Researchers may acquire these through inter-library lending, but there are costs involved in delivery and the process may take days, weeks, or months. Print theses do not easily incorporate emerging digital forms of scholarly communication, such as audio files, video files, electronic data sets, and multimedia. Finally, storing print theses takes up valuable stack space in libraries with limited room for new materials.

Electronic theses and dissertations can address these problems and even go beyond them to provide benefits that are impossible with print theses. In *Issues and Innovations in Electronic Theses and Dissertations*, George J. Soete recognizes the many benefits of ETDs.² Through online repositories, ETDs are accessible "from anywhere at any time." By automating the submission process, ETDs are quickly available after submission. Born-digital documents provide the full-text for indexing, enhancing their "searchability." The inclusion of multimedia enhances the quality of the theses and contributes to the learning process of the student. ETDs have the potential to provide savings in storage, circulation, loan, and processing costs. Through the addition of metadata, ETDs simplify the cataloging process. In general, the process of producing an ETD educates the student in the "use of electronic technologies in scholarly inquiry and publication." Considering that scholarly communication is moving rapidly online,³ ETDs help graduates to be better prepared for future careers in research and teaching.

Virginia Tech first began exploring ETDs in the early 1990's and began accepting ETDs in 1996. They founded the Networked Digital Library of Theses and Dissertations (NDLTD), a membership organization that engages in many activities to promote the growth of ETDs. Since its founding, it has grown into an international organization with over 170 members.⁴

The NDLTD has made great strides by promoting the adoption and creation of ETD systems, but there is still much work to be done. The Texas Digital Library seeks to break the mold of campus-based ETD repositories by developing and implementing the Texas ETD Repository, a state-of-the-art ETD system to manage the full life-cycle of ETDs in the State of Texas. This system will integrate policy, training, software, hardware, and stewardship to manage ETDs from their initial creation and distribution to their long-term preservation.

The Texas Digital Library

The Texas Digital Library (TDL) is in an excellent position to address broad needs for the international ETD community while fulfilling the ETD vision in Texas. TDL's mission is to offer a broad set of scholarly services to the university community, enhancing productivity and enabling a wider distribution and use of the intellectual output of faculty and students.⁵ Established in the summer of 2005, TDL's founding members include the four public Association of Research Libraries (ARL) universities in Texas: Texas A&M University, Texas Tech University, The University of Texas at Austin and the University of Houston. Subsequent members include Angelo State University, Baylor University, Texas State University - San Marcos, the University of North Texas, the University of Texas at Arlington, the University of Texas at Brownsville, the University of Texas at Dallas, and the University of Texas M. D. Anderson Cancer Center. All six of the State of Texas' university systems, which are comprised of 42 degree granting institutions, have representation in TDL's current membership.

Dr. John Leggett of Texas A&M and Mark McFarland of the University of Texas serve as Co-Directors of TDL. Support for operations is provided by a small technical staff and a number of working groups consisting of members drawn from TDL's participating institutions. These Working Groups include Computing Infrastructure, Electronic Theses and Dissertations, Metadata and Cataloging, Repositories, Preservation, and Web Oversight. The University of Texas at Austin is providing facilities for dedicated TDL personnel.

The members of the TDL represent the highest caliber universities in the state of Texas. Considering that the current members of the TDL account for nearly three-quarters of Ph.D.s granted in the state,⁶ TDL has an inherent interest in the success of ETDs.

Assessment of Need

Regionally, there is a significant need for a state-wide ETD initiative in Texas. Although there are thirty four Ph.D. granting institutions in the state, only six universities in Texas are members of the NDLTD. Only three TDL universities (The University of Texas at Austin, Texas A&M University, and Texas Tech University) have implemented ETD systems. If the ETD movement continues to grow organically with the adoption of ETD systems on a campus-by-campus basis, the rate of adoption will be slow and uncoordinated, and could result in multiple incompatible systems. By developing a state-wide plan, the TDL can leverage its resources to develop new and innovative technologies while coordinating the development of an interoperable, state-wide ETD system that can serve as a model for other states and/or countries.

On a national and international scale, there is a need for leaders to develop the next generation of ETD systems, services, and standards. The most widely used system, Virginia Tech's ETD-DB, was developed in December 1999 and is showing its age.⁷ Other systems are being used to manage ETDs, most notably DSpace, but DSpace's workflow is based on a pre-print self-archiving model, not a thesis model. Its inflexible workflows are not tailored for ETD submission. In 2004, Edinburgh University developed a customized DSpace plug-in for ETDs, the Theses Alive Plug-in for Institutional Repositories (TAPIR).⁸ Both TAPIR and ETD-DB are missing key features that are necessary for managing future ETD collections.

The time is ripe for the development of the next-generation ETD system that is capable of managing the entire ETD workflow. Such a system would provide features such as administrative and structural metadata to organize complex multimedia ETDs, a digital rights management system to provide strong protection of intellectual property, advanced bibliographic description, and a robust and redundant digital preservation system. The Texas ETD Repository will address each of these needs in turn by utilizing the Metadata Encoding and Transmission Standard (METS) and its extension schemas⁹, as well as a preservation network infrastructure based on the Texas Internet Grid for Research and Education (TIGRE) and the Lonestar Education And Research Network (LEARN).

National Impact and Intended Results

The Texas Digital Library will have a significant national impact through the development of a state-wide comprehensive ETD repository built upon a widely-accepted open source repository system.

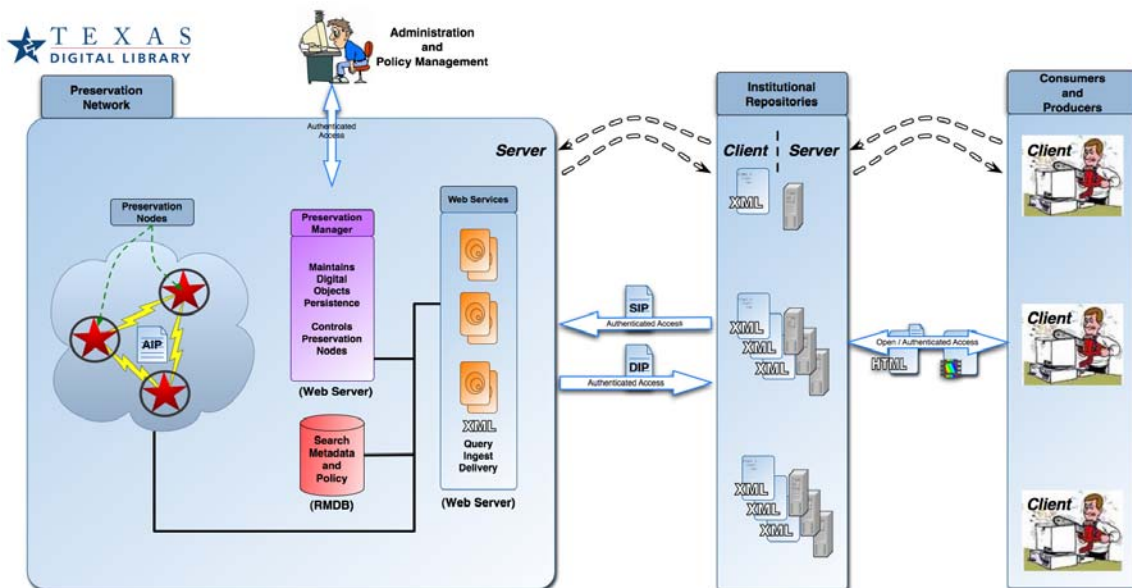
The Texas ETD Repository will use DSpace technology, but will extend its function through the use of Manakin, a framework for DSpace that allows the development of custom user interfaces.¹⁰ Manakin, an open-source project developed at Texas A&M University in collaboration with the DSpace Federation, was released in January 2007. Manakin significantly expands the usability of DSpace by allowing repository managers to make changes to the functionality and display characteristics of a DSpace instance. By replacing the inflexible Java Server Pages User Interface (JSPUI) which DSpace uses with a much more adaptable Extensible Markup Language User Interface (XMLUI), Manakin changes the way a user interacts with a DSpace repository. When content is generated from a DSpace repository in XML rather than JSP, XSL stylesheets can be applied to change the look and feel of what is ultimately presented to the user. Further, repository administrators can develop new features, or “aspects”, for the DSpace repository. This new technology will allow TDL to develop several new features critical to the success of a state-wide ETD repository.

First, Manakin allows for the customization of the DSpace presentation at multiple levels, from top-level collections down to individual items. TDL will build a federated repository to provide a single point of discovery for all ETDs throughout the State of Texas. A search of all Texas ETDs will be possible from this central repository. Further, Manakin will allow each ETD in the repository to remain branded with its academic institution’s logo. Because of this, users will be able to filter the federated repository, allowing them to search for publications originating from a particular institution. The ETD Repository will be Open Archives Initiative (OAI) compliant so that external services, such as scholars’ portals, search engines, or citation linking systems may provide access as well.¹¹ A functional prototype of a Manakin-altered Texas Digital Library repository, which demonstrates its customization ability for DSpace and the branding of individual items, can be seen at <http://repositories.tdl.org>.

Second, using the aspect development feature of Manakin, TDL will develop a comprehensive ETD workflow system that will be used by each of its member institutions. This system will manage three stages of the ETD workflow: ingest, verification, and publication. The ingest stage will gather the ETD itself, any supplemental materials, metadata, and copyright authorizations from the student. Once the complete ETD package is in the system, the Thesis Office at the

student's academic institution will begin the verification process, ensuring that metadata is complete and accurate, and will undergo the iterative proofing process, sending corrections to the student, who makes the corrections and resubmits the ETD. Once the ETD has cleared the Thesis Office, the ETD and its metadata will be assigned to the repository, with considerations for embargoes for publication or national security reasons.

TDL will also coordinate the long-term security of the collection by implementing a state-wide preservation network that will comply with Trusted Digital Repository and OAIS specifications. The goal is to spread the responsibility for managing and maintaining a trusted digital repository over all of the members of the TDL in order to reduce the risk associated with reliance on any one institution's resources. Submission and subsequent access of materials to the preservation network will be made by the institutional repositories, and storage of those items will be made in multiple and geographically diverse locations. Combined with rich preservation metadata based on the PREMIS metadata standard, the preservation network will enable high quality stewardship for these valuable scholarly resources.



Further, TDL will address the shortcomings of Dublin Core by developing and implementing metadata standards based on METS. Maintained by the Library of Congress, METS is a standard for encoding administrative metadata using the XML schema language. METS has the potential to serve the needs of the ETD community through the inclusion of different metadata standards for organizing and describing compound ETDs and encoding rights management and preservation metadata. Through the collaboration of librarians at each member

institution, a descriptive metadata application profile based on MODS has already been developed¹², and efforts are being made to draft metadata standards for rights management, a preservation metadata profile based on the PREMIS metadata standard, and METS-based metadata for the encoding of compound digital objects. Throughout the project, TDL will work directly with the NDLTD to contribute to the development of standards for metadata.

Project Design and Evaluation Plan

The Texas ETD Repository goals are to:

1. Develop policies for ETD submission, licensing, access, and preservation.
2. Develop standards for structural, descriptive, rights management, and preservation metadata using METS.
3. Develop a state-wide common ETD submission system with organic METS support.
4. Develop multi-faceted repository discovery services.
5. Develop a robust digital preservation system.
6. Deploy the system for the members of the Texas Digital Library and recruit other graduate institutions in the state to participate.
7. Sustain and evaluate the Texas ETD Repository.

These goals will be achieved over the next three years:

Year 1 (2007-2008): Planning and Development.

TDL will focus on the development of key components of the ETD Repository in its first year. Key deliverables include:

- The ETD Working Group will develop a common state-wide submission system for ETDs, which offers complete control of the ETDs collected by the system, from initial submission through final publication. This will be accomplished by examining member institutions' current processes and developing and quality testing a usable prototype.
- The Metadata Working Group will have primary responsibilities for developing metadata standards for rights management, preservation, and compound objects, based upon accepted and emerging metadata standards such as MODS, PREMIS, and CopyrightMD, with METS as the base standard for administrative and structural metadata.
- The Computing Infrastructure Working Group will design, develop, and test a state-wide digital preservation system in order to provide long-term stewardship of the repository. This repository will follow

the OASIS model, be geographically distributed, redundant, and have advanced functionality enabled by OAI-PMH and rich preservation metadata.

- Establishment of institutional repositories will be accomplished in this year. TDL will also develop a federated repository that maintains university branding at the item level to provide a single access point for all ETDs in the state. Besides these “direct” user interfaces, each institutional repository will be OAI enabled to allow external services to harvest the metadata. These include commercial web search engines (Google, Yahoo, MSN, etc.), scholarly portals (Google Scholar, Scirus, Current Web Contents, etc.), and reference linking systems (SFX, CrossRef, etc.).

Year 2 (2008-2009): Deployment of the ETD Repository.

In the second year, TDL will deploy the submission, discovery, and preservation systems developed in Year 1. This phase will require a high degree of coordination among the TDL members.

- The common ETD submission system will be implemented on a demonstrative basis for Texas A&M, Texas Tech, and The University of Texas at Austin. After obtaining feedback and testing the system, the submission system will be ready for statewide deployment at the beginning of Year 3. TDL will provide technical support, orientation, and training to each campus.
- The federated repository will be implemented by the TDL to provide one point of access for all of its members’ institutional repositories.
- The preservation network will be implemented. Preservation storage nodes will be installed on the TDL members’ campuses and the institutional repositories will be configured to communicate with the preservation system.
- Following deployment among current TDL members, further graduate degree institutions in the state will be encouraged to participate. Although the current members of the TDL account for 73% of Ph.D.s in the state, coverage will be increased by broadening participation in the Texas ETD Repository to include other graduate degree granting institutions in the state.

Year 3 (2009-2010): Sustain and Evaluate.

After initial deployment, TDL will enter the maintenance period for the repository during which each campus will use the system for ongoing ETD management. The working groups will continue to address emerging

maintenance issues to ensure that the system is satisfying its functional requirements. TDL will evaluate the Texas ETD Repository using multiple methods: user studies, usage statistics, and repository certification. User studies will assess service quality through surveys of faculty, students, and repository users. Usage statistics, based on server log analysis, will be employed to evaluate user access of the repository. This analysis can contribute to an understanding of how the various discovery services lead users to the content in the repository. TDL will also develop a certification process that defines the minimum qualities and services to be provided by TDL members regarding ETDs.

Project Resources: Budget, Personnel, and Management Plan

The Texas A&M Research Foundation will administer the project grant and will handle all accounting and grant management functions, including disbursement of funds. All purchasing activities related to the grant will follow State of Texas purchase and procurement laws and guidelines. Recruitment and hiring are governed by State of Texas personnel policies.

Dr. John Leggett and Mark McFarland will serve as co-Project Directors and will supervise all staff that will have direct responsibility for project activities and deliverables. TDL will recruit and hire a Texas ETD Repository Project Manager who will report to the Project Directors, manage all of the technical aspects of the project by working closely with TDL's Steering Committee and working groups. The Project Manager will be responsible for managing the life of the project and reporting progress to the Project Directors.

Each member of TDL will provide support in the form of participation on the five working groups. These working groups will coordinate their activities to design and deploy the repository throughout the life of the project. Dedicated TDL staff will contribute by participating in the working groups and focusing their efforts on the development of the federated search system and the preservation network.

The major portions of the budget are portions of salary for two individuals at each founding TDL institution: a metadata librarian whose activities would include policy planning for metadata within the ETD repository and coordination with library cataloging and the graduate studies office at their institution in implementing the system; and an application developer whose responsibilities would be developing the common submission system, the institutional interfaces, and the ETD repository's preservation network. Further, these people would assist in the implementation of the system at new member institutions as they join the Texas Digital Library.

Dissemination

The project participants are well positioned to disseminate the outcomes of the project. TDL will use various channels of dissemination: the annual ETD conference, the annual Open Repositories conference, the DSpace Federation User Group Meetings, the bi-annual Digital Library Federation (DLF) Forum, and publication of journal articles. TDL will send representatives to the annual ETD conference to disseminate updates on the Texas ETD Repository through presentations and poster sessions. TDL will also communicate the results of research and development to the NDLTD Standards Committee with the goal of contributing to the development of future standards. Texas A&M is a member of the DSpace Federation and will report on progress at the DSpace Users Group Meeting. The University of Texas at Austin is a member of the DLF and will report on progress at the bi-annual DLF Forum. Finally, participants in the project will publish the results of research in appropriate journals and publications, such as D-Lib Magazine, Ariadne, First Monday, the International Journal on Digital Libraries, etc.

Sustainability

The Texas ETD Repository will be a production-level system to be used for managing all aspects of theses and dissertations. The Repository will fill an existing and ongoing need for the members of TDL. The design of the Repository includes a robust preservation system to address this need. Administratively, TDL will provide the necessary services to sustain the system. The members of TDL have each made significant monetary contributions to fund TDL for the next three years.

The Texas Digital Library has also put into effect a membership fee for all institutions. It is intended that once all aspects of the ETD Repository are built and implemented, those funds will be sufficient for the maintenance and ongoing costs of its operation.

This proposed National Leadership Grant will contribute to the development of next generation scholarly communication tools through the implementation of a state-wide electronic thesis and dissertation system. This project will have significant positive impacts at the campus, state, national and international level, and will expose valuable scholarly resources that have traditionally been underutilized. The Texas ETD Repository promises to provide significant benefits for all stakeholders in graduate level scholarly communication.

References

- ¹ Council of Graduate Schools. 1991. *The Role and Nature of the Doctoral Dissertation: A Policy Statement*. Washington, D.C.: Council of Graduate Schools.
- ² Soete, G. J. 1998. *Issues and Innovations in Electronic Theses and Dissertations*. Washington, D.C.: Association of Research Libraries.
- ³ Odlyzko, A. 2002. The Rapid Evolution of Scholarly Communication. *Learned Publishing*, 15(1), 17-19.
- ⁴ Networked Digital Library of Theses and Dissertations. 2005. *NDLTD Members*. Retrieved December 20, 2005, from <http://www.ndltd.org/membership/dir.html>
- ⁵ Texas Digital Library. 2005. *Texas Digital Library: A Business Case*. Internal memorandum.
- ⁶ Hoffer, T. B., V. Welch, Jr., K. Williams, M. Hess, K. Webber, B. Lisek, D. Loew, and I. Guzman-Barron. 2005. *Doctorate Recipients from United States Universities: Summary Report 2004*. Chicago: National Opinion Research Center.
- ⁷ French, L. 2004. *Release History for the ETD-db Distribution*. Retrieved December 20, 2005 from http://scholar.lib.vt.edu/ETD-db/developer/download/RELEASE_HISTORY
- ⁸ Jones, R. 2004. The TAPIR: Adding E-Theses Functionality to DSpace. *Ariadne*, 41, Retrieved December 20, 2005, from <http://www.ariadne.ac.uk/issue41/jones/>
- ⁹ Library of Congress. 2005. *METS: Metadata Encoding and Transmission Standard*. Retrieved December 20, 2005, from <http://www.loc.gov/standards/mets/>
- ¹⁰ Phillips, S., C. Green, J. Leggett, A. Maslov, A Mikeal, and B. Surratt. 2005. *Manakin Developer's Guide*. Retrieved December 20, 2005, from <http://di.tamu.edu/projects/xmlui/manakin/resources/DevelopersGuide.pdf>
- ¹¹ Open Archives Initiative. 2004. *The Open Archives Initiative Protocol for Metadata Harvesting*. Retrieved December 20, 2005, from <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- ¹² 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