

EDUCAUSE Center for Applied Research

Research Bulletin

Volume 2006, Issue 13

June 20, 2006

# Institutional Strategies and Policies for Electronic Theses and Dissertations

Joan K. Lippincott, Coalition for Networked Information



## Overview

Almost without exception, students produce theses and dissertations in electronic formats. It would seem that an institutional electronic thesis and dissertation (ETD) program, in which an institution mounted the culminating work of its graduate students on a publicly accessible Web site, would be the rule and not the exception. In the United States, however, ETD programs have had a relatively slow uptake; other countries are far ahead in implementing comprehensive strategies and frameworks for the creation of, access to, and stewardship of ETDs.<sup>1</sup> The seemingly obvious and straightforward notion of putting a collection of institutional intellectual output on the Web turns out to be a project that requires some changes in institutional policies and practices as well as the support of a wide range of stakeholders across campus. So, like many other issues involving technology, the technical aspects of the project are straightforward, but the institutional and policy issues create barriers to progress.

ETD programs provide an opportunity for the institution or its departments to consider allowing students to enrich the content of their final products with nontextual digital materials that allow for a fuller expression of students' ideas or a clearer representation of their work. These digital formats could include such things as simulations of engineering processes, 3D models of chemical molecules, audio files of oral history interviews, and videotaped dance performances.

Theses and dissertations freely available on the Internet are globally accessible documents that represent the scholarship produced at an institution. They raise the profiles of the students who author them, the faculty and departments who foster them, and the institutions that provide them to the world.

This research bulletin provides an overview of the rationale for an ETD program, describes the ETD process, and explores institutional policy issues. It is intended to assist graduate school deans and other administrators, CIOs, faculty serving as advisors, and directors of libraries and digital library programs at institutions in the early planning stages of ETD programs by identifying key issues and providing resources on model programs.<sup>2</sup>

## Highlights of ETD Programs

Traditionally, master's and doctoral students have followed strict institutional procedures to produce print copies of their theses or dissertations for submission to their review committee, the graduate school, or other administrative entity, and for archiving by the institution's library or outside entities. Today, most theses and dissertations are produced as files using word-processing software but are still submitted to the institution as print copies. In a full-service ETD program, a range of processes related to the thesis or dissertation are automated, including

- certification by the committee,
- submission to the graduate school or other administrative unit,

- processing by the library or other systems for cataloging or metadata creation or verification,
- access to the full content, and
- preservation of the content and format.

Basically, an ETD program provides a process, standards, and software that automates functions as well as a digital infrastructure for access to and preservation of the content of the ETD.

Some institutions that have an ETD program require that all students participate and do not accept print submissions; other institutions keep participation voluntary. Some institutions accept ETDs but have not automated the submission or certification processes; in other words, students submit paper copies to their committee and to the university administration, and another entity such as the library or an outside vendor accepts and maintains the digital copy.

In the early 1990s, some pioneers were already envisioning the potential of theses and dissertations as drivers for digital libraries. ETDs were seen as low-hanging fruit, lacking the complex intellectual property issues surrounding the digitization of commercially published scholarly products. Early adopting institutions such as Virginia Tech have had ETD programs in place since the mid-1990s. The first wave of ETDs was generally stored as a part of a digital library collection administered by the library. In recent years, an ETD program has frequently served as the foundation or pilot for institutional repository content as institutions begin to develop a broad framework and infrastructure for the stewardship of the digital intellectual assets produced by members of their community. In addition, ProQuest, the commercial company that has been the central source for collections of dissertations and the provider of index and abstract services that assist individuals in locating dissertations of interest, has embraced electronic submission, access, output, and preservation. Other commercial entities such as the Online Computer Library Center (OCLC) and Visionary Technology in Library Solutions (VTLS) have entered the ETD marketplace with technical solutions.

Importantly, ETDs are an institutional issue, not the solution of an individual acting alone. If a student puts his or her dissertation on a Web site, it is a digital dissertation but not part of an ETD program. An institutional ETD program identifies and preserves the authoritative copy of record of the student's work. In contrast, there is no guarantee that the content of an independent digital document will survive into the future, no institutional commitment to stewardship (migration to current software and updated platforms), and no protection of the integrity of the content (against intentional or unintentional changes).

The Networked Digital Library of Theses and Dissertations (NDLTD), a global organization committed to the acceleration of institutional adoption of ETD programs, has participation from institutions on six continents. It disseminates information on such topics as standards, metadata, preservation, student and staff training, and policy issues related to ETDs through its Web site, publications, and annual conference.

An institution should take the following considerations into account when establishing an ETD process on campus.

## Involvement

A range of individuals and units within a university are involved in the establishment of an ETD program and its operation. These include graduate school administrators, academic administrators (provosts, deans, department heads), faculty advisors, students, librarians, and information technologists.

- Typically, the graduate school in a university sets policy for the formal process of submission and approval of theses and dissertations, the acceptable format(s) of theses and dissertations, and the manner in which they are managed in the institution.<sup>3</sup> They provide the institution's final stamp of approval that the candidate has met the thesis or dissertation requirements for a degree.
- Other academic administrators may participate in the decisions made by the graduate school, or they may set more specialized guidelines or policies for students in particular programs. For example, in the performing arts or sciences, departments may want to specify format standards that are particularly useful or acceptable in their disciplines.
- Faculty thesis and dissertation advisors help determine the acceptability of the type of content that students may incorporate into their final product (for example, use of nontext formats) and will likely be asked by their students to serve as a resource on the ETD submission process. Even when an institution has an electronic submission process in place, faculty members who do not wish to receive digital copies of the thesis in its various stages or at the final stage may request that a student print out a copy for his or her review.
- Students, of course, are the creators and authors of the ETDs and need to understand the process used by their institution and the implications for the availability of their intellectual work.
- Librarians manage the processing of the digital content into the library and generally provide metadata (cataloging) associated with the document. They provide an infrastructure, whether it be a digital library infrastructure, an institutional repository, or an arrangement with an outside vendor, to allow access to the full text of the dissertation. They handle licensing for their own campus access to the ProQuest index and abstract service (Dissertation Abstracts) and to dissertations distributed by that company. In addition, they are either stewards for the long-term preservation of theses and dissertations or they advise the institution on the mechanisms for outsourcing that function.
- Information technologists, along with librarians, may provide training for both faculty and students to use ETD software and to learn how to format their content to acceptable technical standards. Information technologists also provide the network infrastructure needed to support both the creation of ETD

content and the submission of final products. They are responsible for network security issues, and they may advise on format standards.

When an institution begins considering implementing an ETD program, typically an administrative unit such as the graduate school will bring together a committee or task force of representatives from these various groups to discuss procedures and policy issues.

## Submission Software

A number of software products are available for the ETD submission process. Virginia Tech's software is freely available and adoptable or adaptable by others. In 2005, VTLS released VALET, a free, open source, Web-based solution for ETD submission into a Fedora repository. Recently, UMI/BEPRESS developed an online submission application for ETDs that was used as part of a Council of Graduate Schools study.

## Format Standards

An ETD program articulates document format standards and institution-defined policies. These standards take into account traditional text format documents and also address other digital formats. Many U.S. colleges and universities with ETD programs are currently using Adobe PDF as the standard document type format. PDF can incorporate a variety of media formats. With funding from Adobe, The Ohio State University developed a tutorial for creating an ETD using Adobe Acrobat that is freely available (see <<http://etd.vt.edu/etdtutorials/>>). Some institutions also have provision for acceptance of ETDs in LaTeX (used primarily in the sciences), and some are exploring XML for the future.

Most institutions with ETD programs specify what types of nontextual materials they will permit in ETDs and which standards should be used for those materials (for example, for video or audio files). Standard formats for ETD metadata, which assist with the retrieval of the digital documents, are used by most institutions with ETD programs. Many ETDs are cataloged using the ETD-MS metadata format, which is based on the Dublin Core. Submission software may include templates for metadata capture in standard formats.

The emergence of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) defines a standard method of extracting metadata from a repository (Open Archives, 2002). This enables the construction of aggregators that can build union databases of ETDs held by consortia members or NDLTD member institutions. Both OCLC and VTLS, with additional international languages capabilities, take advantage of OAI-PMH to promote access to online theses and dissertations (see OCLC Electronic Theses and Dissertations <<http://www.oclc.org/research/projects/etd/default.htm>> and VTLS and NDLTD Scan Search <<http://zippo.vtls.com/cgi-bin/ndltd/chameleon>>).

## Access

Institutions have a wide range of choices about how they wish to approach access by others to the full content of ETDs and how they want to provide for long-term preservation. If an institution hosts its own ETDs, it may do so through its digital library

program, institutional repository software and infrastructure, or some other type of content management system. A variety of factors may influence the mechanisms used by a particular institution, including already available software, staff knowledge, and the overall structure of the digital content in the institution or the library. Some institutions may not directly host their ETDs, choosing consortial or commercial solutions instead. For example, OhioLink hosts ETDs from Ohio institutions, including Ohio State University. Since the late 1990s, ProQuest has provided digital copies of dissertations even if they were submitted in paper form by the student. When determining how the full content of their ETDs will be hosted, institutions should consider

- local versus consortial versus commercial storage;
- software choice, including integration with other digital content; and
- whether it is freely available or offered for a fee to the general public.

Traditionally, individuals located dissertations of interest via the print publication or Dissertation Abstracts online database or through a search of an institution's online catalog. Individual items in those resources are not exposed to Web search engine crawlers. Now, there is a wide array of possibilities for discovering the availability of a thesis or dissertation on a particular topic, by a specific author, or from a department within an institution. These include performing a search using a widely available Internet search engine such as Google, using catalogs that aggregate metadata such as those developed by OCLC and VTLS, and searching Amazon.com. Not all ETDs can be located by each method; factors affecting accessibility include whether the metadata for the ETD is harvestable and whether a particular resource aggregates data from a particular institution.

## Long-Term Preservation

Higher education libraries usually are responsible for the long-term preservation of theses and dissertations produced by students. A variety of options are available for preservation. Some make a commitment to local long-term preservation, particularly if the ETDs are part of an institutional repository. Some libraries are participating in a pilot using the Stanford-developed LOCKSS process for providing long-term access.<sup>4</sup> Some institutions rely on ProQuest to maintain microfilm and digital copies for the long term; others use vendors such as OCLC. While some universities may choose an open access solution for current access to the content of their ETDs, they may choose either a local or a commercial solution to protect the long-term availability of that content.

## Costs

The Virginia Tech library provides some figures on the costs of implementing an ETD program, but these will vary by institution, both in terms of local costs for items like staff salaries and for the extent to which the ETD program is implemented locally (Virginia Tech, 2005). Institutions may save staffing costs in the graduate school if individuals no longer need to be available to physically accept print copies and library costs by reducing binding, shelving, and circulating resources if paper copies are no longer submitted. Students may also realize cost savings if they don't have to purchase

expensive paper and make copies for storage, and if they need not be physically present to submit the final copy of their dissertation, especially if they have left the area. Costs the institution may incur include staff and student training, software, computer storage, and the ongoing maintenance of the digital collection.

## Key Policy Issues

As institutions determine the viability and details of ETD programs, several key policy issues arise.

### Public Availability

While there are many reasons that the broad public availability of ETDs is advantageous, in certain situations an ETD author and his or her advisor may have concerns about the immediate availability of an ETD on the Internet. These situations often involve ETDs that contain materials for which the author will be seeking a patent. Most institutions create policies to deal with this type of situation. They permit student authors to “embargo” access to the dissertation, usually by limiting access for a defined period of time, after which it will be publicly available, or by limiting access to campus network users only.

The issue of public availability also arises if an institution wants to retrospectively scan dissertations produced prior to the existence of an ETD program. Institutions that have begun this type of program often have limited access to the digital copies for campus use: they seldom secured permission from the author to transfer the thesis or dissertation content to new formats when the document was first produced, and it is time-consuming and difficult to track down former students to ask for permission years after the completion of their work.

MIT has taken a unique approach to public access and now has 10,000 selected ETDs available from the period 1879 to the present. Anyone can read the PDF files online through MIT’s DSpace repository, but the ETDs may not be freely downloaded or printed by the general public (MIT Libraries, 2002).

Another dimension to public availability is a new relationship between ProQuest and Amazon.com. Amazon is now serving as an additional source for dissertations held by ProQuest whose authors have not placed restrictions on the distribution of their work. Individuals can use Amazon.com to purchase copies of some dissertations held by ProQuest, and authors receive royalties from ProQuest on the same basis as if they had been purchased directly from ProQuest.

### Prior Publication

Faculty advisors may encourage some students to publish the results of their thesis or dissertation research in established scholarly communication channels, such as books and journals. When discussions are initiated on campus to develop an institutional ETD program, faculty can be vocal opponents—the “prior publication” issue is often paramount in their objections. They are concerned that if their graduates’ works appear or are “published” on a university Web site as part of an ETD program, a society or

commercial publisher may decline to publish their work as a result of this prior publication.

In many disciplines where this issue is raised, it is important to realize that dissertations have a very different structure and format; considerable revision is required to publish them in traditional channels. In some of the sciences, however, the dissertation itself is a compilation of several articles that have either been accepted for publication in a journal or are already published. In these cases in particular, valid concerns may be raised about the potential conflict of publishing the same material on a publicly available institutional Web site as in a journal.

Because this has been a key issue in the adoption of ETD programs and for the success of the open access movement to provide freely available copies of preprints or e-prints of journal articles, some individuals and institutions have developed resources to specifically address the concerns of ETD authors who may want to publish in the future in other outlets on the topic of their dissertation.<sup>5</sup> Interestingly, the University of Tennessee encourages its students to use statistics on access to dissertations online as a means of demonstrating to a potential publisher the market for a book they would create based on the content of the dissertation (University of Tennessee, Knoxville, 2003).

### Plagiarism

Some faculty advisors and others express concern about institutional adoption of an ETD program because they fear that it will open the door to plagiarism of their students' works. This reservation about ETDs seems to be a red herring. Plagiarism of dissertations occurred prior to their availability on the Internet, and now that plagiarism detection software is widely available, it should mitigate these concerns.

### Security

Campus IT experts will want to provide advice on security issues related to an ETD program. These include protecting any dissertations that are embargoed and safeguarding the integrity of documents. Administrators will want to address issues of whether it is advisable to have committee signature pages available in digital form and whether any identifying information that individuals would prefer to keep off the Internet appears in the ETD.

## What It Means to Higher Education

It is increasingly evident that publications that have freely accessible versions on the Web are more widely accessed than those that are available only in print form or via licensed or password-protected access.

### Institutional Visibility

Institutions and individual authors who want wide dissemination of their scholarship will find that freely available Web access of their publications will provide the opportunity for

individuals around the world to benefit from their scholarship. In the past, many academics were unconcerned that dissertations typically languished on library shelves, since they were considered inferior forms of scholarship. Common wisdom was that if the content had academic merit, it would be published later in the more standard print monograph or journal format, after a peer review process. These publications were usually derived from the dissertation content; they were edited, reworked, and revised to suit the standards of the publishing outlet.

As institutions such as Virginia Tech began to make the access statistics of their ETD collections available, others could see that theses and dissertations may have been underestimated as useful resources. For example, one of the most heavily accessed ETDs at Virginia Tech between 1996 and 2003 was produced by an electrical and computer engineering student on a topic related to handheld wireless devices; that dissertation's PDF file was accessed more than 60,000 times. Virginia Tech's dissertations are accessed by individuals coming in from a wide variety of Internet domains, including .edu, .com, and .gov. West Virginia University, which requires ETDs, promotes the program by telling students that their "graduate research documents are now...accessed millions of times per year by academia, industry, government and the public from over 100 countries" (West Virginia University, 2005).

Statistics gathered by Virginia Tech illustrate the huge increase in use of dissertations that are freely available in digital form compared to their use when they were exclusively available in print form. Prior to the availability of ETDs, the average for Virginia Tech's print dissertation circulation was less than one time per item (Virginia Tech, 1998). As institutions, particularly those that are state-funded, make the case to governing bodies and legislatures as to the value of their programs to business and society, this kind of data can provide useful documentation of the wide reach of research produced on campus and the contribution of its programs.

### Educating Authors for the Digital Age

One of the guiding principles of NDLTD is that it is important for "graduate students to learn about electronic publishing and digital libraries, applying that knowledge as they engage in their research and build and submit their own ETD" (Networked Digital Library of Theses and Dissertations, n.d.). Whether university and college students are heading to academic, business, government, or social service careers when they complete their programs, they will likely be producing electronic content in their work and in their personal lives. They need to understand both technical and policy issues related to electronic publishing. An ETD program can be much more than a collection of digital documents—it can be the training ground for digital publication. Students can use the ETD program to gain an understanding of issues related to standardized formats for documents and media. The program can also help them develop an awareness that the amount of time they want a digital object to continue to be available should influence both the format and "home" of that object. In addition, students need to have at least an elementary understanding of intellectual property issues in the context of the use of others' materials in their work and the future use of their dissertation content by others. In a new twist, James Hilton reported on a situation at University of Michigan where an extended negotiation was needed between the student author, faculty advisor, and

university administrators (and attorneys) to determine the copyright terms of a multimedia thesis that a student created in partnership with a faculty member using university-paid videographers (Hilton, 2006).

Educating future authors does not happen by osmosis. Institutions interested in developing digital authors need to commit resources to this outcome. Faculty can play a key role in helping students understand the electronic publishing landscape in their discipline, and librarians and information technology experts can provide training and support for students to learn about software, format standards, and the legal principles of information use and dissemination.

## New Forms of Scholarship

At present, ETDs represent incremental changes in standard, print-based theses and dissertations rather than transformational products of scholarly communication. They are generally text-based and conservative in genre. The creative kinds of products that can be generated in the digital era such as simulations, video, and 3D visualizations are largely absent from ETDs. They represent efforts by students to conform to the requirements of their faculty advisors, committees, departments, and universities. Those in authority do not yet generally accept new media as bona fide outputs of scholarship.

Increasingly, our graduate students will come from the Net Gen student cohort—those students who grew up with computers from the earliest age. They have the potential to develop creative, useful forms of scholarly communication in their disciplines, but they need the encouragement and permission of their faculties and administration.<sup>6</sup> Net Gen students frequently work in groups, much like many professionals who have careers in business or the sciences. Whether institutions will permit students to submit group-authored ETDs in the future remains to be seen.

Even if the body of the ETD is largely text, students may wish to enhance their ETDs with appendices that incorporate their raw research data, databases, spreadsheets, interviews (video, audio), maps, and image collections. Whether in the body of the ETD or an appendix, these media and formats have storage implications, migration issues, proprietary software issues, intellectual property issues, and privacy issues. Complex issues will be attached to new media or new formats in ETDs, and institutions must address them in a way that makes sense in terms of their institutional policies and culture.

## Key Questions to Ask

- Who are the key stakeholders that need to make policy decisions about our institution's ETD program?
- Which units will administer various aspects of the ETD program?
- What components will our institution include in our ETD program?
- Will we encourage creative use of new media in ETDs or take a conservative approach?

- How will we design our ETD program to be a vehicle for educating students to become authors in the digital age, assisting them in understanding technical, presentation, availability, and policy issues related to intellectual output?

## Where to Learn More

- Moxley, J. M. (2001). Universities should require electronic theses and dissertations. *EDUCAUSE Quarterly* 24(3), 61–63. Retrieved March 24, 2006, from <http://www.educause.edu/LibraryDetailPage/666?ID=EQM0139>
- Networked Digital Library of Theses and Dissertations, <http://www.ndltd.org/>
- Networked Digital Library of Theses and Dissertations. (2001, December 6). The guide for electronic theses and dissertations. NDLTD. Retrieved March 24, 2006, from <http://www.ndltd.org/etdguide/ETDGuide.pdf>

## Acknowledgments

The author thanks Gail McMillan at Virginia Tech, Joan Cheverie at Georgetown University, and Clifford Lynch of the Coalition for Networked Information for their suggestions and review of this bulletin.

## References

- Hilton, J. (2006). The future for higher education: Sunrise or perfect storm? *EDUCAUSE Review* 41(2), 58–71. Retrieved March 24, 2006, from <http://www.educause.edu/LibraryDetailPage/666?ID=ERM0623>
- MIT Libraries. (2002). DSpace. Retrieved March 24, 2006, from <https://dspace.mit.edu/index.jsp>
- Networked Digital Library of Theses and Dissertations. (n.d.). History, description, and scope: Goals. Retrieved March 24, 2006, from <http://www.ndltd.org/info/description.en.html>
- Open Archives. (2002, June 14). The Open Archives Initiative protocol for metadata harvesting. Retrieved March 24, 2006, from <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- University of Tennessee, Knoxville. (2003). UT Thesis/Dissertation Website. Retrieved March 24, 2006, from <http://web.utk.edu/~thesis/>
- Virginia Tech. (1998, February 25). Dissertations circulating from University Libraries, Virginia Tech, as of Feb. 1996 and Feb. 1998. Retrieved March 24, 2006, from <http://scholar.lib.vt.edu/theses/data/Dcircs98.pdf>
- Virginia Tech. (2005, August 23). Supporting ETDs at a typical university library. Retrieved March 24, 2006, from <http://scholar.lib.vt.edu/theses/data/setup.html>

- West Virginia University. (December 19, 2005). Electronic theses and dissertations. Retrieved March 24, 2006, from <http://www.wvu.edu/~thesis/>

## Endnotes

1. For example, Canada and Australia have national ETD repositories.
2. This bulletin focuses on graduate theses (master's level) and dissertations (doctoral level). Some institutions may be interested in a similar program for undergraduate theses or senior projects, and many of the issues covered here should apply equally.
3. In some universities, the administration of theses and dissertations may be handled in a decentralized manner, for example by departments or colleges within the university or in a mixed situation, where the department or college handles some functions and the graduate school handles others.
4. LOCKSS, for Lots of Copies Keep Stuff Safe, is open source software that provides librarians with an easy and inexpensive way to collect, store, preserve, and provide access to their own, local copy of authorized content they purchase. Retrieved March 24, 2006, from <http://www.lockss.org/about/about.htm>.
5. The University of Cincinnati's Academic Journal Policy Database (<http://www.etsd.uc.edu/journal/>) focuses on concerns of ETDs, and SHERPA publisher copyright policies & self-archiving list, a resource of the University of Nottingham in the United Kingdom, provides information of use to authors interested in open access publishing in general (<http://www.sherpa.ac.uk/romeo.php>).
6. The theme of ETDs and Net Gen students is more fully developed in Joan K. Lippincott, "Net Gen Students and ETDs," paper presented at ETD 2005, Sydney, Australia. Retrieved March 24, 2006, from <http://adt.caul.edu.au/etd2005/papers/107Lippincott.pdf>.

## About the Author

*Joan K. Lippincott ([joan@cni.org](mailto:joan@cni.org)) is Associate Executive Director at the Coalition for Networked Information.*

Copyright 2006 EDUCAUSE and Joan K. Lippincott. All rights reserved. This ECAR research bulletin is proprietary and intended for use only by subscribers. Reproduction, or distribution of ECAR research bulletins to those not formally affiliated with the subscribing organization, is strictly prohibited unless prior permission is granted by EDUCAUSE and the author.