THESES AND DISSERTATIONS FOR THE NEXT MILLENNIUM

Janet Webster
Guin Library, Hatfield Marine Science Center
Oregon State University
2030 Marine Science Drive
Newport, OR 97365 US

Abstract: All educational institutions granting higher degrees have the responsibility to archive copies of their dissertations and theses. While some people prefer the peer-reviewed articles that should emerge from the work, others find value in the originals. This debate aside, dissertations and theses present an intriguing opportunity for electronic archiving and retrieval. Is electronic storage feasible, economical, and reliable? What are the copyright considerations? Will it change how and what students submit? What does a librarian need to consider before moving towards electronic storage? The technology is developing and librarians need to consider how to use it.

Keywords: Copyright; Dissertations, Academic; Electronic Publications; Intellectual Property

Introduction:

An underlying principle of scholarship is that original work should be disseminated to the community. Theses and dissertations represent the original work of graduate students. University libraries traditionally archive and provide access to this work, thus encouraging dissemination (DesJardins 1998). With the advent of electronic publishing, some institutions are exploring electronic theses and dissertations (ETDs) as a means of improving access as well as efficiently archiving them. The following overview provides background on ETDs, a brief discussion of the issues, and advice for institutions considering implementing them.

Why worry about theses and dissertations?

Why should we even consider new ways to handle theses in our institutions? Three arguments bolster the case for focusing attention on theses: increased volume in the number created, evolving perceptions of value, and problematic access.

The rate of graduate degrees conferred in many disciplines is increasing. Consequently, the growing volume pressures on physical space and technical services workflow. Some estimate that 40,000 doctoral degrees and 360,000 masters degrees are awarded in the

United States annually resulting in approximately 100,000 theses and dissertations (Fox et al. 1996). *Science and Engineering Indicators 1998* shows steady output over the last 20 years in the United State for degrees in the natural sciences (see Table 1). In 1997/98, Oregon State University conferred 232 doctoral degrees and 669 masters degrees; the OSU Libraries added over 500 theses and dissertations with 1997 and 1998 publication dates. University graduate school offices and libraries must handle this output.

Table 1 Graduate Degrees Granted in the United States

	1975	1985	1995
Physical Sciences	3,076	2,934	3,840
Atmospheric, Oceanographic & Earth Sciences	625	599	778
Biological & Agricultural Sciences	4,402	4,903	6,406
TOTAL	**8,103**	**8,436**	**11,024**

Over the years, theses and dissertations have been both denigrated and lauded. To some, they are "the ultimate homework" (Fox 1985) while others view them as "the first work of the scholar" (Rutledge 1994). Studies reveal that theses are used, and hence valued, in varying degrees according to discipline and type of user. A recent British survey of recent doctoral students, their major professors, and librarians found that 86% consulted a thesis or dissertation in the course of their work (Roberts 1997). Rutledge summarizes many usage studies with a range of observations; one citation study found scanty use of dissertations in the hard sciences while another noted a rising level of dissertations cited in the geological literature (1994). Usage varies by discipline, with some using theses extensively – library science and education, for example. Primary users appear to be graduate students (Repp & Galviano 1987; Lee-Smeltzer & Hackleman 1995). However, staff members at Virginia Polytechnic Institute and State University (Virginia Tech) have found a wider variety of users of ETDs including occasional heavy use by commercial research and development organizations (Fox *et al.* 1998).

The question of usage is a question of access: "They aren't made adequately accessible because use is thought to be low, and, in practice, they won't be used because they aren't easily accessible" (Rutledge 1994, p.56). Currently, enhancing access to theses and dissertations beyond borrowing the original includes indexing them in broadly used resources and publishing the material in either journal or book form. Few masters theses are indexed broadly, so access is very limited. Doctoral dissertations fair better as many are indexed in *Dissertation Abstracts*.

The academic myth that the good ones get published is not necessarily so. An interesting 1975 survey showed that one third of American academicians never published (Ladd & Lipset 1975). Lopez (1988) looked at the top five universities in Political Science and Sociology, surveyed authors of dissertations, and found that only 50-60% of these "good" dissertations were published as books and articles. Others have found similar results with variance by discipline -- some of the hard sciences having a higher success rate than humanities and social sciences (Lopez 1988). Personal experience reveals that not all good, original work gets published-- particularly masters level work in aquaculture and marine mammal science. Reasons vary from lack of interest on the student's part, lack of follow-though by the major professor, or lack of outlets.

So, as the volume of theses and dissertations ratchets up and shelf space is consumed, storage and access appear to be problematic. Granted, the situation has improved with greater indexing and better coverage in the electronic databases, although an update of Lopez's and Hartman's 1988 study could be done to verify this. Online university catalogues make it possible to locate theses. Even if once located, however, they can be difficult to borrow. These factors have led several universities and their libraries to examine ETDs.

Why consider electronic theses and dissertations?

ETDs are an intriguing means of archiving and making graduate students' work accessible. They have several strengths.

• Enhance access through open access as well as more complete cataloging.
ETDs are easily searchable if located and linked correctly. They can be accessed both locally and remotely via the Web. Usage at Virginia Tech has increased dramatically since making theses and dissertations web-accessible (Scholarly Communications Project 1998).

• Allow experimentation with digital collections.
Theses and dissertations as a discrete part of our collections provide an opportunity to move towards digital collections. They are manageable; they come into our collections in a proscribed way; and we have sole institutional responsibility to care for them. They are a potential test digital collection.

• Encourage students to publish electronically.
Graduate student education is evolving as is scholarly communication. Students need to be well-versed in electronic information. That means not only knowing how to access it, but also how to create it. ETDs are a means of codifying electronic publication.

• Integrate a wider range of formats and information.
ETDs can have more dimensions than print ones. They can be less linear and more expressive including images, sound, and data files. For example, a physical

oceanography student could include her computer model of eddies in full color and motion.

• Save time.

Graduate students would save time in submitting their work. At Virginia Tech, students submit via the Web, attaching the ETD and associated files. Consider the student who had to reformat and reprint her thesis at least twice due to a persnickety Department Chair. That meant four trips to the main Oregon State University campus or almost 500 miles of traveling-- and a good deal of ranting and raving. If it had been an ETD, the file could have been sent back and forth, reviewed, and corrections attached. Technical services would save time in cataloging and processing. Users would find theses and dissertations in a more timely fashion. For example, York University and the University of Toronto found that it currently took, on average, one year for a print thesis to get through the entire technical process and on the shelf ready for use (DesJardins 1997a). An electronic one could be opened and used soon after being accepted by the graduate school.

• Save shelf space.

Instead of more shelving and additional archival quality storage, we would add another file server or more electronic storage.

• Save money.

Saving graduate students money for copying and binding is a plus. Saving libraries money is perhaps more elusive.

What are ETDs?

Technically, ETDs are electronic files in a standard format archived on a server and accessible via an electronic gateway. The formats currently in vogue are Portable Document Format (PDF) and Standard Generalized Markup Language (SGML). Adobe Acrobat, a readily available PDF reader, allows a document to retain its formatting and graphics, facilitates indexing, and is easy to use (Weisser et al. 1997). PDF serves as an "electronic wrapper." The drawbacks of Adobe Acrobat are its proprietary nature and the limitations of searching and indexing.

Standard Generalized Markup Language is an evolving standard for web based documents. SGML is "an internal standard for the definition of device-independent, system-independent methods of representing texts in electronic form" (Sperberg-McQueen & Burnard 1998). It relies on the use of a set of tags that serves as the "grammar" for a given document and is developed so that it can be submitted and converted to a format that preserves its formatting while lending itself to electronic access.

Including images, video, and audio allows the content to expand from traditional text and figures to a myriad of options. Standard formats for these are emerging (Fox et al. 1998; DesJardins1997b). Once the file is created and submitted to the institution, it is stored on a central file server and a gateway is created for users.

What is the history and status of ETDs?

In 1987, a meeting at University Microfilms (UMI) led Virginia Tech to help fund development of an SGML format. Since then, Virginia Tech has pursued ETDs energetically, working with the Southeastern Universities Research Association and Southeastern Libraries Network. In 1996, the U.S. Department of Education funded the creation of the National Digital Library of Electronic Theses and Dissertations (now known as the Networked Digital Library of Electronic Theses and Dissertations). The Library had 36 members worldwide as of July 1998 and is very pro-active in encouraging others to join (Networked Digital Library 1998).

While Virginia Tech is the most visible institution in the field, UMI has launched its Digital Dissertation Project to better support scholarly communication (and perhaps its bottom line.) Since 1997, UMI scans all theses and dissertations it receives and creates a PDF file. It currently has well over 60,000 titles digitally available and linked through Dissertations Abstracts (Fox 1998). This number compares to 1200 for Virginia Tech.

Table 2: Useful Resources

Joint Electronic Thesis & Dissertation Project	http://www.fis.utoronto.ca/etd	good overviews and discussion of the issues
Networked Digital Library of Electronic Theses & Dissertations	http://www.ndltd.org/	practical help and links to relevant material
UMI Digital Dissertations	http://wwwinfo.umi.com/solutions/	explanation and demonstration of commercial endeavor
University of Texas at Austin/ Report of the Ad Hoc Committee	http://www.utexas.edu/orgs/organizations/dissertations.html	example of university working group recommendations

What are the issues for librarians?

It is one thing to read about what others are doing, and another to do it. The issues to consider surface consistently and can be grouped around three processes: getting ETDs into the collection, keeping them in the collection, and making them accessible. Some elements are straightforward; others are not. Some only involve the library while others require cooperation and coordination across campus.

Getting ETDs into the collection

Getting ETDs into the collection requires that a student create it before the library can receive, catalog, and process it. More and more, libraries are becoming involved in helping students create their work. For example, the Guin Library at the Hatfield Marine

Science Center in Newport, Oregon houses the student computer lab and at times, the staff feels inadequate in dealing with software questions. Librarians are skilled in teaching about citations and search strategies, but often need to make some alliances with others to teach strategies and skills for the creation of electronic documents. The equipment and expertise should be accessible to all students. So, both can be located in departments, computer labs or libraries. Oregon State University's expanded main campus library will house an information commons complete with equipment to access and manipulate electronic information. There, students will find assistance from student workers and support staff in the creation of electronic documents. Librarians will have to work with others on campus to ensure that students are trained and that standard formats are adopted that are simple to use.

Once created, the library has to receive, catalog and process the theses. Will this be done by a simple file transfer or will staff members need to login to a controlled system? At Virginia Tech, once the thesis is reviewed and accepted by the Graduate School, it goes to the library collection. There, the cataloguer logs in to the ETD system and creates a MARC record from the information form submitted by the student (McMillan 1996). This form can provide extensive information including keywords, notes on appendices and illustrations, and even an abstract. Such information would greatly enhance most cataloging records, and keep theses and dissertations from languishing while librarians catch up with the original work needed (Lee-Smeltzer& Hackleman 1995). Virginia Tech estimates processing costs of $3.20 per electronic volume versus $12 for a print one (DesJardins 1997b).

What about the appendices and problems? Using the Virginia Tech model, the Graduate School is responsible for reviewing the ETD. It is probably the appropriate place for review as the student is still around. If it is done by the library, the student will have graduated and moved on, and will not be available to provide a missing piece or explain a non-standard format.

The ETDs move out of the cataloguer's file to a permanent machine. That machine needs to have sufficient storage and be easily accessible. Most ETDs currently require about one megabyte of storage. Those with images will require substantially more -- 5-10 megabytes (Fox et al. 1996).

Keeping ETDs in the collection

Nobody has a perfect solution to preservation and storage. Virginia Tech talks about a three part plan that includes maintaining standards, conducting multiple and frequent backups, and refreshing when necessary (DesJardins 1997b). UMI maintains two servers, one for access and one for archiving. It also still keeps multiple formats -- electronic, CDs, print and fiche.

Tackling the preservation issue means deciding who is responsible for what. The student is one place to start. In *Preserving Digital Information*, creators of digital information objects are responsible for the creations (Task Force 1996). While fine in theory, students rely on the University and its library to archive their theses. The Library has traditionally been responsible. However, if the server housing the ETDs is located outside the library and maintained by non-library staff, agreements must be arranged to assure the safe archiving of the ETDs. Another option for archiving is an outside agency or consortium. UMI serves as something of an archive and central clearinghouse in the United States and has been adding some Canadian dissertations since 1990 (Olson 1995). The Networked Digital Library is a repository for its members and has storage for an estimated 40 million ETDs. The National Library of Canada is the recognized repository for Canadian theses and dissertations. Some combination of local and remote archiving is a possibility to consider.

Once the ETDs reside somewhere in some kind of standard format, what happens to them in ten years?

There are two very different approaches: refreshing and migrating. Refreshing means copying them 'as is' to a new machine and assumes that the format is still readable. This may be fine if the original standard is still a standard. Migrating means converting to a new standard or an updated version. Migration is potentially much more expensive but is the right thing to do. The Task Force on Preserving Digital Information found that "as long as the preservation community lacks more robust and cost-effective migration strategies, printing to paper or film and presenting flat files will remain the preferred method of storage for many institutions" (Task Force 1996). Exon suggests that "the best chance electronic information has of being preserved is that it should go on being used, regularly and continually (1998, p.3). In other words, electronic information is not robust and libraries must create effective strategies if ETDs will be accessible in years to come.

The cost is another intriguing aspect of the storage issue. Reliable figures are scarce. People at Virginia Tech estimate that disk space for 1000 ETDs would cost less that $3,000 annually (Fox et al. 1996). That can add up over time. The Task Force on Digital Preservation compared print to digital depositories using a model developed by Yale Library. Those same 1000 ETDs would cost $210 annually to shelve in storage. Even expanding that to less compact shelving, it still is significantly less than digital. However, while digital storage is more expensive, digital access is cheaper. Enhancing access must be the main reason for considering ETDs.

Making ETDs Accessible

Onsite access involves supplying adequate equipment for viewing, the right software, and printing capabilities. The equipment and software will follow from the decisions made about format standards and preservation approach. A potential pitfall is the specter of generations of technology, something many libraries already have to deal with. They do

not seek out additional layers. Another pitfall is the cost of printing out the ETDs. Many libraries are absorbing growing printing costs as information moves to electronic format. Libraries need to consider whether savings from binding will go towards printing on demand or if a pay-to-print system should be implemented.

Access for both local and remote users demands a usable gateway with a good search engine. The online catalog with links in the 856 field would provide unified local access. Additional links or web gateways are probably needed to take full advantage of the broader search possibilities.

Increased electronic access can bring security problems. It is imperative that ETDs remain unchanged once submitted and accepted. The locked vault in Special Collections seems like a pretty simple security solution compared to the complexities of electronic document security. Levels of access may be needed with passwords for those needing to do file maintenance. Security must be discussed as part of long-term storage and maintenance plans.

What about information ethics?

So far, the issues are technical: format, maintenance, processing workflow. What about issues of information ethics: copyright, academic publishing, and plagiarism? These issues cause many students to look the other way or when forced to listen, their eyes glaze over. These issues are crucial when considering ETDs as they press students to decide about copyright, free access, and academic communication.

These entities (ETDs) will be in library collections because they have been released by their creators for the use of others. The copyright resides with the creator, with the student. One exception is if the student did the work as a paid contractor (the usual graduate student assistantship does not count.) Occasionally, a thesis leads to a patent; that patent is almost always owned by the university; but the copyright on the publication is still the student's.

Students sign agreements about distribution and they should question what they are signing. At OSU, the release is for use by library readers. Virginia Tech has a non-exclusive license to archive and make theses available in all forms of media. UMI negotiates for the exclusive right to distribute copies in or from a microfilm copy, and a non-exclusive right to archive and make copies on demand. These layers of licenses can be contradictory, confusing and potentially harmful to a student's growth as a scholar. A license to distribute in all forms does not automatically restrict access; if an ETD is on the web, there is the possibility that anybody can get it. Virginia Tech students were concerned about the effect that wide-open access to their work would have on their ability to publish (Young 1998). Publishers opinions are mixed on whether they consider ETDs prior publication (Guernsey & Kiernan 1998). Virginia Tech has instituted levels of access to alleviate the concerns. Theses range from full access to no access until the

student releases the ETD. A student should be encouraged to investigate the state of her discipline and its publication policies before deciding whether to publish electronically. The publishing field is constantly shifting and it is difficult to track the changes.

Discussing ETDs provides a venue to debate open and free access to theses and dissertations versus a student's right to publish for reputation and money. If some ETDs are widely popular, should that student get some financial reward? Or should she be content that she has contributed to the growth of knowledge? At UMI, if a work is purchased seven times or more in a calendar year, that author gets royalties (Savage, 1996). It happens rarely. Yet, some ETDs at Virginia Tech have been accessed 9,000 times (Fox et al 1998). Is the information free through the library or should users pay? Many faculty members are concerned with plagiarism. ETDs do not change the concern one way or another, except that it is easier to plagiarize by cutting and pasting out of an electronic document than a print one. This is not a format question, but rather an ethical one. Are librarians and faculty members explaining the nuances of licensing? Are institutions of higher education doing a good job teaching the ethics of scholarship, from copyright to plagiarism? These questions will have to be addressed before implementing ETDs.

Are ETDs in your future?

Before committing your institution, look at what others have done. Virginia Tech has experience and is willing to share it. The technical issues alone are complex, let alone the philosophical ones.

Decide if ETDs make sense for your institution and its disciplines. There is variety in how researchers in different disciplines use and value theses and dissertations. Surveying in-house use and inter-library loans would provide helpful information (Lee-Smeltzer & Hackleman 1995). A starting point would be to focus on a select number of disciplines or departments that are important to your institution's core mission. If your institution and its people do not place value on these works, then perhaps it would be more logical to put effort into other parts of your collection.

If there are indications of interest in better access to theses and dissertations, broaden the discussion of ETDs to include the faculty, graduate school, and students. It is critical that the faculty and students share an attitude that supports exploring electronic publishing. They also need to have the skills to create the publications. So, a campus-wide assessment of skill levels would be important. Eventually, a task force or committee would take the information gathered, synthesize it and make recommendations on whether to implement ETDs. This would lead to development of policies and guideline for creation and archiving, drafting an implementation strategy, and deciding who is responsible for teaching how to create ETDs.

All of this requires institutional commitment. ETDs are intriguing both technically and educationally. Mere interest in the technical aspects will not be enough to shape graduate student scholarship in the next millennium. Institutional commitment to electronic scholarly communication is the key. This means money, because any major shift to ETDs takes some capital. If Virginia Tech is a model, it takes a lot. It also means a clear understanding of the interplay of student's rights, open access to information, and scholarly communication. An institution that provides adequate access, supports the creation of electronic publications with equipment and expertise, and teaches the ethics of information can tackle the question of utilizing ETDs. Without such commitment, ETDs will be just another technical comet and an opportunity to explore new ways of promoting graduate student work will fade.

References:

DesJardins, J. 1998. "ETDs and Publishers." [http://www.fis.utoronto.ca/etd/report4.htm].

DesJardins, J. 1997. "Paper Thesis and Dissertation Workflow Report." [http://www.fis.utoronto.ca/etd/report2.htm].

DesJardins, J. 1997. "Status of ETD Initiatives in the US and Canada." [http://www.fis.utoronto.ca/etd/report1.html].

Exon, M. 1998. Long-Term management issues in the preservation of electronic information. In: *1995 National Preservation Office (NPO) Conference: Multimedia Preservation: Capturing The Rainbow* (ed. by M. Exon), pp. ??? National Library of Australia, Brisbane, Australia.

Fox, E.A., Eaton, J.L., McMillan, G., Kipp N.A., Weiss, L., Arce, E. & Guyer, S. 1996. "National Digital Library of Theses and Dissertations: a scalable and sustainable approach to unlock university resources." [http://www.dlib.org/september96/theses/09fox.html].

Fox, E.A., McMillan, G. & Eaton, J.L. 1998. "The evolving genre of electronic theses and dissertations." [http://www.ndltd.org/pubs/Genre.htm]

Fox, M.F. 1985. The transition from dissertation student to publishing scholar and professional. In: *Scholarly Writing and Publishing: Issues, Problems and Solutions* (ed. by M.F. Fox), pp.6-16. Westview Press, Boulder, Colorado.

Guernsey, L.& Kiernan, V. 1998. Journals differ on whether to publish articles that have appeared on the web. *The Chronicle of Higher Education* July 17:A27-A29.

Ladd, E.C. & Lipset, S.M.. 1975. How professors spend their time: the 1975 Ladd-Lipset survey of U.S. faculty members. *The Chronicle of Higher Education* 11:2.

Lee-Smeltzer, J. & Hackleman, D. 1995. Access to OSU theses and dissertations in Kerr Library: how are they used... or are they? *Technical Services Quarterly* 12(4):25-43.

Lopez, M.D. 1988. Dissertations: a need for new approaches to acquisition. *The Journal of Academic Librarianship* 14(5):297-301.

McMillan, G. 1996. Electronic theses and dissertations: merging perspectives. *Cataloging & Classification Quarterly* 22(3/4):105-25.

Networked Digital Library of Theses and Dissertations. 1998. "Information about the Initiative." [http://www.ndltd.org/].

Olson, P. 1995. Obtaining translations and academic theses: an international survey. In: *Building Global Connections: Proceedings of the 20th Annual Conference of the International Association of Aquatic and Marine Science Libraries and Information Centers* (ed. by J.W. Markham & A.L. Duda), pp.123-161, IAMSLIC, Fort Pierce, Florida..

Repp, J.M. & Glaviano, C. 1987. Dissertations: a study of the scholar's approach. *College & Research Libraries* 48(March):148-59.

Roberts, A. 1997. "Survey on the Use of Doctoral Theses in British Universities." [http://www.;ob.ed.ac.uk/Theses/index.htm].

Rutledge, J.B. 1994. European dissertations: production, access, and use. *Collection Management* 19(1/2):43-67.

Savage, W.A. 1996. *Publishing your Master's Thesis: How to Prepare your Manuscript for Publication.* UMI Dissertation Services, Ann Arbor, MI.

Scholarly Communications Project. 1998. "Virginia Tech electronic theses and dissertations: facts, data, information: ETDs at University Libraries, Virginia Tech." [http://scholar.lib.vt.edu/theses/data/datahp.htm.].

National Science Board. 1998. "Science & Engineering – 1998." [http://www.nsf.gov/sbe/srs/seind98/].

Sperberg-McQueen, C.M. & Burnard, L. 1998. "A Gentle Introduction to SGML: Chapter Two of Guidelines for Electronic Text Encoding and Interchange." [http://etext.virginia.edu/TEI.html].

Task Force on Archiving Digital Information. 1996. "Preserving Digital Information: Report of the Task Force on Archiving Digital Information ." [http://www.rlg.org/ArchTF/].

Weisser, C., Baker, J. & Walker, J.R. 1997. "Problems and Possibilities of Electronic Theses and Dissertations." [http://www.december.com/mag/1997/nov/reds.html].

Young, J. R. 1998. Requiring theses in digital form: the first year at Virginia Tech. *The Chronicle of Higher Education* Feb 13:A29.