

# Cover Sheet: FIPSE Final Report

Grantee Organization:

Office of Sponsored Programs  
Burruss Hall  
Virginia Polytechnic Institute and State University  
Blacksburg, VA 24061

Grant Number:

P116-B-61190-96

Project Dates:

Starting date: September 1, 1996  
Ending date: August 31, 1999  
Number of months: 36

Project Title:

**Improving Graduate Education with the  
National Digital Library of Theses and Dissertations**

Project Director:

Edward A. Fox  
Professor, Department of Computer Science  
phone: +1 540 231 5113  
fax: +1 540 231 6075  
email: fox@vt.edu

FIPSE Program Officer:

Donald Fischer

Grant Award:

Year 1	69,762
Year 2	69,337
Year 3	68,941
Total	\$208,040

## Improving Graduate Education with the National Digital Library of Theses and Dissertations

With FIPSE support, university efforts improving graduate education through electronic theses and dissertations (ETDs) spread nationwide, helping thousands of students gain knowledge and skills with electronic publishing and digital libraries. Led by the Virginia Tech project team, this national project became an international initiative with more than 70 members, the Networked Digital Library of Theses and Dissertations, with popular WWW sites at [www.ndltd.org](http://www.ndltd.org) and [www.theses.org](http://www.theses.org), supported by UNESCO and Organization of American States. Cooperative, funded, national projects began in Australia and Germany. ETDs represent a new genre for scholarly communication, with hypermedia content, read thousands of times more than paper theses.

### Project Director:

Edward A. Fox  
 Professor, Department of Computer Science  
 Virginia Polytechnic Institute and State University  
 Blacksburg, VA 24061  
 phone: +1 540 231 5113  
 fax: +1 540 231 6075  
 email: [fox@vt.edu](mailto:fox@vt.edu)

<p><b>Web sites:</b>  <a href="http://www.ndltd.org/">http://www.ndltd.org/</a>  <a href="http://www.theses.org/">http://www.theses.org/</a>  <a href="http://etd.vt.edu/">http://etd.vt.edu/</a></p> <p><b>Primary publications:</b></p> <p>Rekha Kengeri, Cheryl D. Seals, Hope D. Harley, Himabindu P. Reddy, Edward A. Fox. "Interface and evaluating: Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD" <i>International Journal of Digital Libraries</i>, September 15, 1999, 2:157-169.</p> <p>Edward Fox, "Networked Digital Library of Theses and Dissertations," <i>Nature, Web Matters</i>, August 12, 1999, <a href="http://helix.nature.com/webmatters/library/library.html">http://helix.nature.com/webmatters/library/library.html</a></p> <p>Constantinos Phanouriou, Neill Kipp, Ohm Sornil, Paul Mather, and Edward A. Fox, "A Digital Library for Authors: Recent Progress of the Networked Digital Library of Theses and Dissertations," in <i>Proc. ACM Digital Libraries '99</i>, Berkeley, CA, Aug. 11-14, 1999.</p> <p>E. Fox, G. McMillan, J. Eaton, "The Evolving Genre of Electronic Theses and Dissertations", in Proceedings Digital Documents Track of HICSS-32, Thirty-second Annual Hawaii Intl Conf. on Systems Sciences, Maui, HI - January 5-8, 1999.</p> <p>George J. Soete, ed. (Gail McMillan, Ed. Advisor), "Transforming Libraries 7: Issues and Innovations in Electronic Theses and Dissertations," Association of</p>	<p>Research Libraries, October 1998, ISSN 0160-3582, <a href="http://www.arl.org/transform/index.html">http://www.arl.org/transform/index.html</a></p> <p>James Powell and Edward A. Fox. "Multilingual Federated Searching Across Heterogeneous Collections," <i>D-Lib Magazine</i>, September 1998, <a href="http://www.dlib.org/dlib/september98/powell/09powell.html">http://www.dlib.org/dlib/september98/powell/09powell.html</a></p> <p>John L. Eaton, Edward A. Fox, Gail McMillan, "The Role of Electronic Theses and Dissertations in Graduate Education," <i>CGS Communicator</i>, Jan. 1998, <a href="http://www.cgsnet.org/Communicator_Online/Jan98/Lead198.html">http://www.cgsnet.org/Communicator_Online/Jan98/Lead198.html</a></p> <p>Neill A. Kipp, "Case Study: Digital Libraries with a Spatial Metaphor," in <i>SGML/XML '97 Conference Proceedings</i>, Graphic Communications Association, December 1997.</p> <p>Edward A. Fox, John L. Eaton, Gail McMillan, Neill A. Kipp, Paul Mather, Tim McGoingle, William Schweiker, and Brian Devane. "Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources," <i>D-Lib Magazine</i>, September 1997, <a href="http://www.dlib.org/dlib/september97/theses/09fox.html">http://www.dlib.org/dlib/september97/theses/09fox.html</a></p> <p>Edward A. Fox, John L. Eaton, Gail McMillan, Neill A. Kipp, Laura Weiss, Emilio Arce, and Scott Guyer. "National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources," <i>D-Lib Magazine</i>, September 1996, <a href="http://www.dlib.org/dlib/september96/theses/09fox.html">http://www.dlib.org/dlib/september96/theses/09fox.html</a></p>
---	--

# Improving Graduate Education with the National Digital Library of Theses and Dissertations

## Executive Summary

### *Project Overview*

FIPSE support has enabled a small pilot effort, focused on enhancing graduate education through digital library technology, to become an international initiative demonstrating the potential for university collaboration and federation. Launched as the National Digital Library of Theses and Dissertations in September 1996, NDLTD was renamed in spring 1997 as the Networked Digital Library of Theses and Dissertations, due to international interest. Though this report focuses on US activities, our project has played a crucial role in helping Australia and Germany launch similar funded efforts, has guided UNESCO in its efforts to establish a global initiative, has helped the Organization of American States and the Ibero-American Science and Technology Education Consortium develop a hemispheric program, and has assisted consortia as well as individual universities in other countries including Canada, China, India, Italy, Mexico, Portugal, Russia, Singapore, South Africa, South Korea, Spain, and UK. These international efforts have direct benefit to students in the US, who have little access to theses from other nations; a collaboration has developed with the Center for Research Libraries, in Illinois, that provides access to paper theses from outside North America - in addition to many other project partners.

Prior to the end of 1999, over 70 organizations, almost all universities, have officially joined NDLTD. More than one-half are in the US, spread around the nation. This reflects the widespread and ongoing adoption of the goals of the initiative and the change in practice at universities to enhance graduate education.

### *Purpose*

This project aims to enhance graduate education through a digital library of theses and dissertations. The ultimate aim is for every undergraduate or master's thesis, and all doctoral dissertations, to be produced in an electronic form and placed in a digital library by the author. This should ensure that the next generation of scholars, and key leaders of our society as well as academia, have at least basic "information literacy" that includes knowledge and skills in word processing, electronic publishing, and digital libraries. It should encourage every university to have at least a base level of computing, networking, and multimedia technology so that their students can most effectively communicate new research findings through electronic documents. Because graduate students are among the most frequent users of others' theses and dissertations, this effort should encourage more study of the relevant literature by graduate students and other researchers who wish to find in-depth descriptions of scholarly investigations. Since theses and dissertations are one of the richest sources of literature reviews, having a digital library of theses and dissertations should make it easier for those interested in a field to gain an overview of related works.

By engaging each member university in the process of establishing a program for electronic theses and dissertations, this effort serves as a seed project in the areas of electronic publishing and digital libraries. It aims to heighten awareness of the process of scholarly communication, the connections between universities and outside organizations (e.g., publishers, associations), the economics of and legal issues related to intellectual property, methods of markup, requirements for metadata, urgent needs for preservation, and the expanding opportunities for international collaboration. Through involvement in the federated searching service of NDLTD, and the discussions at workshops and conferences, it promotes sharing and collaboration among universities to provide better service in a time of scarce resources.

## ***Background and Origins***

The idea of electronic theses and dissertations (ETDs) was first discussed publicly in a workshop hosted by University Microfilms International (UMI) in 1987 that involved Virginia Tech and other parties. Practical solutions became feasible in the early 1990s with the arrival of Adobe's Acrobat software, which supported creation of Portable Document Format (PDF) documents. Workshops and meetings extended the discussion, with additional help by the Council of Graduate Schools and the Coalition for Networked Information. By the end of 1995, the Southeastern Universities Research Association had agreed to fund a regional effort, led by Virginia Tech, building tools, infrastructure, and training materials, as well as disseminating workable techniques to interested universities. FIPSE funding, for the three-year period starting with September 1996, has spread the effort across the nation and encouraged activities around the globe.

## ***Project Description***

A large part of this effort has involved dissemination. The project team has given hundreds of presentations, explaining possibilities and practice at individual universities, often through several visits. This process continues, with trips requested, at the expense of the interested universities, through late in 2000. Many active NDLTD members have visited Virginia Tech, sometimes for several days, to obtain a deeper understanding of the issues, approaches, technologies, policies, and potentials. The project team has presented at regional, national, and international conferences as well as to focused groups representing particular communities, such as graduate deans and university librarians. Consortia have been briefed, including the CIC (for the "Big 10" universities) and OhioLink (for the public and private universities in Ohio). Larger gatherings have brought interested parties together in annual meetings, held in Memphis, Tennessee in 1998 and in Blacksburg, Virginia in 1999. A major international conference on ETDs is planned for Florida in 2000, with thousands of invitations sent. It has support by diverse parties including the Council of Graduate Schools, UNESCO, and Microsoft.

Information about NDLTD is available from three Web sites, each with a different target audience:

- the project site, at <http://www.ndltd.org/>,
- the collection site, at <http://www.theses.org/>, and
- the students' help site, at <http://etd.vt.edu/>.

## ***Evaluation/Project Results***

Our careful and involved evaluation of the project demonstrates that we have achieved the goal of improving graduate education. Students are learning more about electronic publishing. Technology is being transferred to others. Students and universities are saving time and money. We demonstrate all this through numerous log analyses of Web sites, surveys of students and digital library patrons, comparative cost analyses, participant interviews, and by studies with focus groups.

**NDLTD.** From our involvement and analysis of the general aspects of the NDLTD, we drew the following conclusions:

- The NDLTD is a community-type digital library. Most of our efforts directly support the community of those interested in ETDs.
- We have more than 70 members representing more than a dozen countries, with 20-25 members joining per year.
- We have visited (or have received visitors from more) than 30 of the United States.

- ETD patrons are using the NDTLD Federated Search System.
- For electronic submission to be accepted everywhere, supporting technologies must be integrated at every work point and submission must be mandated.
- Face-to-face meetings are necessary to help potential members understand issues they will face to make ETDs a requirement at their institutions.
- Of our members West Virginia University has already implemented mandatory ETD submission. They have already collected 400 ETDs. NCSU does not have mandatory submission, and has only 72 ETDs. MIT is scanning TDs into electronic form (on demand) and has 3500 available for purchase.

### **ETDs at Virginia Tech.**

- ETDs are getting accessed thousands of times per day, for a fraction of the per-access cost of their paper cousins. Virginia Tech's 2000 ETDs are being accessed as many times per week as 9000 TDs were in five years.
- The mean ETD size is growing each year (from 1.5 megabytes to 3 megabytes) with the additional use of rich multimedia.
- Of surveys filled in, university researchers that are already familiar with PDF are the majority of ETD users.
- On average, students had NOT yet, but planned to publish articles in peer-reviewed journals (70%); did NOT plan to publish a book or were not sure they would (80%); if publishing, planned to publish within 12 months (2/3); if withholding worldwide access, were doing so because of planned publication of parts or whole (3/4); and were not willing or not sure that they would need the ability to release only specific chapters of their ETDs for worldwide access (90%).
- Workshop attendance has leveled off at about 30 students, and workshop questions are about specific applications of technology rather than about rationale for the initiative.

### **Summary and Conclusions**

During the three years of this project, there has been a coordinated program of local development and testing, active dissemination, and careful evaluation. Methods for handling electronic theses and dissertations have been refined, exported, adapted, and further improved to reduce labor and save costs. Educational materials have been prepared, used, and revised in accordance with survey results and ongoing changes in related technologies. The work at Virginia Tech has been adapted to the needs of universities that have begun investigations or pilots, or have risen to the level of joining NDLTD.

More than 70 institutions have joined NDLTD. Students have produced thousands of theses and dissertations in electronic forms. Dissertations sent to Bell and Howell Information Systems (formerly UMI) since 1997 are scanned and made available electronically (for a fee) as well. Clearly, electronic theses and dissertations are becoming the way of the future.

Evaluation has helped improve the project tools and services as well as approaches and methods. The increase in interest in theses and dissertations that results from affording electronic access has motivated universities to reap these same benefits as they enhance the visibility of their students. Data on savings through shifting from paper documents demonstrates additional benefits to both students and universities.

As interest in NDLTD continues to spread, and services increase and expand, we hope that there will be even further benefit to graduate education, university research, and the scholarly community in general.

# Improving Graduate Education with the National Digital Library of Theses and Dissertations

## ***A. Project Overview***

FIPSE support has enabled a small pilot effort, focused on enhancing graduate education through digital library technology, to become an international initiative demonstrating the potential for university collaboration and federation. Launched as the National Digital Library of Theses and Dissertations in September 1996, ND LTD was renamed in Spring 1997 as the Networked Digital Library of Theses and Dissertations, due to international interest. Though this report focuses on US activities, our project has played a crucial role in helping Australia and Germany launch similar funded efforts, has guided UNESCO in its efforts to establish a global initiative, has helped the Organization of American States and the Ibero-American Science and Technology Education Consortium develop a hemispheric program, and has assisted consortia as well as individual universities in other countries including Canada, China, India, Italy, Mexico, Russia, Singapore, South Africa, South Korea, Spain, and UK. These international efforts have direct benefit to students in the US, who have little access to theses from other nations; a collaboration has developed with the Center for Research Libraries, in Illinois, that provides access to paper theses from outside North America in addition to many other project partners.

Prior to the end of 1999, over 70 organizations, almost all universities, have officially joined ND LTD. More than one-half are in the US, spread around the nation. This reflects the widespread and ongoing adoption of the goals of the initiative and the change in practice at universities to enhance graduate education.

## ***B. Purpose***

The purpose of this project is to improve graduate education by starting a national program of electronic submission of theses and dissertations. This should reduce costs at institutions because electronic document handling can be less expensive. A key desired result is a rapidly growing, sustainable, distributed digital library of theses and dissertations, based at each of the Nation's universities, so access to the newest research can be inexpensive and easy.

The primary problem to solve is that graduate students generally are not information literate. Many students are duplicating the work of other graduates, often because it is too difficult and too expensive to find and peruse recent theses and dissertations (TDs) from other institutions. Furthermore, while the electronic age is already upon us, most graduate students have not yet learned how to publish electronically.

Libraries must be efficient with their spending especially because of cost increases in journal subscriptions and resulting budget shortfalls. While libraries spend millions per year (sometimes buying back research results from their own universities!), a significant number of journal articles are outcomes of students' thesis or dissertation efforts. Libraries also must spend money to share TDs through interlibrary loan. Further, there are long-term losses in the paper archive of stolen copies as well as wear and tear on a non-renewable resource such as a hardbound thesis.

Another problem is that many graduate programs have not kept up with technology. The resulting opportunity costs (if universities were later to convert paper documents to high-quality electronic forms) indicate they are losing up to \$30 million per year by inaction! Why have so few universities developed local efforts to support an infrastructure of knowledge sharing?

Finally, in 1996, there was low acceptance of electronic publications for scholarly purposes, particularly in the promotion and tenure process. Why should choice of media affect one's career process? After all, the bulk of journal articles are prepared first by electronic means—using a word processor. Why not take advantage of documents "born electronic" and let them contain the richness that the new medium can afford without having to "dumb them down" to paper and then store them on black and white microfilm?

Therefore, by this project, we are enhancing graduate education by supporting electronic submission, encouraging better and faster knowledge transfer and reuse, meanwhile saving institutions and students money and bringing universities fully into the Information Age.

### **C. Background and Origins**

Virginia Tech started planning its Electronic Thesis and Dissertation (ETD) initiative in 1987. Early participants in ETD meetings included representatives from the University Library, Graduate School, Computing Center, and the Department of Computer Science. Virginia Tech ran pilot studies in 1994, and hosted a workshop sponsored by the Southeastern Universities Research Association (SURA). Virginia Tech began accepting ETDs in 1995. Late in 1995, SURA agreed to contribute funds in 1996 to launch an effort in the Southeast. The Faculty Development Initiative, which trains all Virginia Tech faculty members in advanced computer, communications, and educational technologies, began training faculty to help with ETDs in summer 1996. Virginia Tech's Commission on Graduate Studies and Policies established in Spring 1996 the requirement that all TDs submitted after 1996 be in electronic form. In September 1996, FIPSE funding began.

### **D. Project Description**

As the project progressed, four centers of activity emerged (discussed in the first four subsections following, D.1–D.4). Our efforts have been divided equally among them:

- help student authors preparing ETDs,
- help graduate schools and libraries receiving ETDs,
- build and support ETD repositories and access services, and
- nurture a community of interest in electronic scholarly publishing.

#### **D.1 ETD Authors**

The Networked Digital Library of Theses and Dissertations (NDLTD) is a digital library (DL) for authors (Phanouriou, *et al.*, 1999). In this project component, we developed training materials, held workshops, wrote and improved submission software and prepared assistance packages for students using PDF, LaTeX, SGML, and XML.

We learned that while students have a high opinion of their abilities, many did not have adequate word processing skills for developing ETDs. This led us to concentrate on creating tutorials for building fundamental word processing skills. We also encouraged use of new features in PDF files, helped with specific tools for image processing, and built support for use of markup languages.

**Training materials.** Our training materials (<http://etd.vt.edu/howto/>) include hints for successful ETD development and PDF conversion. Producing a PDF file is a relatively simple process, but making it look professional is much more difficult. Early in the grant period we learned that authors were having significant problems with PDF, especially as they were creating PDF files from a variety of word

processing platforms. We prepared step-by-step instructions for students that teach them how to use their word processors correctly to:

- create tables of contents,
- number pages and sections,
- identify paragraphs and paragraph breaks, and
- format bibliographies and block quotes.

Once the word processor skills were learned, students had a much easier time "distilling" the document into a professional-looking PDF file. (The alternative tool, Adobe PDFWriter, though still packaged with the Acrobat software, is known by us and Adobe to be deficient when processing complex documents such as ETDs. Using PDFWriter can lead to anomalous results when PDF documents are viewed on other machines.)

Additionally, we developed training materials for ETD submission. In the early days, submission required some knowledge of HTML (which exceeded the average students' base knowledge). Furthermore, we found that students did not know how to set up metadata fields correctly.

Once the training materials were prepared, we held campus-wide workshops for ETD development and submission. It was not enough to show students how to use Acrobat "live" (plus, most of the computer classrooms on campus did not have the Distiller installed), so we made Web-based slide shows of the conversion process, which have important advantages:

- we can hold workshops in any Internet classroom (worldwide),
- students can access the slide shows to learn from and for reference (worldwide), and
- students can use these online materials in their jobs after graduation (and they do).

Although the live workshops were very popular (attended by 30-40 students per offering), many students requested videotape recordings of the workshops. This was unexpected. We therefore taped our ETD workshops. These tapes circulate through the University Libraries and through the Graduate School. They are particularly popular with students who are working through the Virginia Tech Northern Virginia Graduate Center. Copies of these tapes have been used at a number of universities around the nation.

**ETD Submission Wizard.** Our first efforts at supporting the act of electronic submission were FTP-based. Web-based CGI scripts immediately replaced FTP as the most popular way to submit ETDs. These scripts were adapted from the library's electronic course reserve project. After using them, however, we found them to be awkward to use, especially when the Graduate School requested students to correct and resubmit ETDs (or, more likely, particular chapters). Therefore during 1998 we redesigned the ETD receiving system with ease of use in mind, following the "wizard" graphical interface metaphor. Now students can set up an "ETD account" and maintain its contents during the last few months of the thesis process. This system also is useful as a backup system for cases where a student's ETD is lost due to computer or disk failure.

**ETD-ML.** The Electronic Thesis and Dissertation Markup Language (ETD-ML) was developed as the forward-looking Standard Generalized Markup Language (SGML) system for ETDs. The first prototype system of ETD-ML was developed prior to FIPSE involvement in this initiative. When ETD-ML (beta version) was debuted, it met with some interest and had some early adopters, but it has not yet become a popular alternative to PDF. Several reasons explain its slow acceptance. Foremost, students are used to using Word or Word Perfect and prefer working in those environments; meanwhile tools that support SGML are rare and expensive, and are too general to support the specific and widely varying needs of ETD developers. Even so, we developed and maintain ETD-ML tools for ETD publishing (as a Web site of document type definitions, Perl scripts, training materials, and documentation; <http://etd.vt.edu/etd-ml/>).

Furthermore, even before XML became a full "recommendation" of the World Wide Web Consortium (<http://www.w3.org/XML/>), we developed support for XML, which has further shifted interest away from our SGML version of ETD-ML. In the coming years, we predict students' common desktop tools will be XML-ready (the latest version of Corel Word Perfect shows definite promise), whereupon we can encourage markup-based (e.g., XML) submissions.

**LaTeX.** To support our (significant minority of) LaTeX users, we immediately developed a LaTeX template that conforms to the Virginia Tech Graduate School's recommendations for electronic submission, and provided training materials so that students who wanted to use LaTeX could do so with minimum impact (<http://etd.vt.edu/latex/>). The resulting ETD is easier to manage and looks more professional. Once LaTeX is mastered, it will serve the student throughout his or her career, especially in engineering, mathematics, or the computing sciences.

However, there is a significant challenge in bringing the output of LaTeX (PostScript, but where the text usually is presented as 300dpi bitmaps, rather than as smoothly drawn characters that are useful at any resolution) to PDF. We set up a special service for students called DVIPDF that does the LaTeX to PDF conversion automatically. The technology uses the same "submission wizard" idea to accept LaTeX submissions, convert them into DVI, then produce PostScript (with "smooth" fonts), and finally use a server version of Adobe Distiller to convert them into PDF. This site is extremely popular on campus among LaTeX users; it is used as much for chapters of ETDs as it is for journal papers, fliers, and homework assignments. We have documented the process so that individual sites working with LaTeX can switch easily to use the "smooth" fonts too.

## D.2 ETD-DL Patrons

Also important to this initiative is the site development for the ETD repository—a digital library in its own right. Here graduate students and others can browse the ETDs already submitted or search for text within the ETD corpus. They can use directly the search service offered by a particular institution, or multicast their searches to each participating ETD library using the NDLTD federated search system (Powell and Fox, 1998). Furthermore, to maintain the aesthetics of library presentation, we have prototyped a visual browser for the Virginia Tech collection that presents the ETDs as books on shelves in a virtual library using VRML (Kipp, 1997).

**VT-ETD site development.** The technical staff at Virginia Tech's Digital Library and Archives department is responsible for maintaining the local library of ETDs. Along with their efforts to smooth the ETD submission process and make the system more usable, members of this staff have simultaneously improved the ETD browsing process. Now we have improved capability for ETD browsing and searching: ETDs are sorted by author's name and ETDs are sorted by department name, with the department being chosen from a controlled vocabulary. Before this update, the access was only by author. More importantly, however, the metadata database is constructed so that a browsing interface could be presented along each axis of metadata; as we obtain a critical mass, particularly useful might be orderings by committee chair and date of defense. All browsing interfaces (Web pages) are updated nightly.

Software for ETD site development is written in Perl and uses the MySQL database tool. This software and its corresponding documentation is packaged periodically and made available without charge to other NDLTD institutions (from <ftp://www.ndltd.org/>).

In Fall 1999, we added Z39.50 support to the Virginia Tech ETD collection. The collection is now searchable by a Z39.50 client, as now we have set up and maintain a Z39.50 server. Left is to provide a Web interface that formats the Z39.50 results into a more readable format—effectively applying a style sheet to the result set.

**Federated Search.** Following the need for a single point of contact to every NDLTD member's ETD collection, but keeping with the general desire at each site to maintain their own collection, we developed a federated search system (available from <http://www.theses.org/>) that multicasts queries to NDLTD member sites. The system consults site configuration files to determine how to transliterate search arguments for the various federated systems. Results for each site can be visited one by one (Powell and Fox, 1997).

To build this service, designers focused on simple queries and created a new application that mediates search requests. This solution homogenizes the variations between search engine user interfaces and query interfaces. It also has the advantage of allowing additional manipulation of queries to improve retrieval effectiveness. This middle tier of software has access to descriptions of the search engines' user interfaces, the types of queries supported, and the operators that define and qualify those queries. Each site's description includes information about operators for formulating a query and about user interface primitives (e.g., buttons, text fields, menus) necessary for the interaction. Because Web searches tend to be only 1-2 words in length, this system focuses on this single input field approach.

The query and user interface description language designed to be used for the configuration files is the Searchable Database Markup Language (SearchDB-ML, an application of XML)

**3DL.** 3DL presents the ETD collection as a three-dimensional VRML model through which users can navigate. It mimics a traditional library: including lobbies, elevators, doors, signs, displays, windows, artwork, doors, rooms, bookcases, and books. Doors are hyperlinks to rooms and books are labeled hyperlinks to items in the ETD collection. In addition to the usual library components, 3DL uses images extracted from the collection and presents them as hyperlinks in a "virtual art gallery." VT-ETD developed 3DL as an alternative interface to the ETD collection (Kipp, 1997).

**CAVE-ETD.** CAVE-ETD extends the 3DL project from the desktop to an immersive virtual reality environment. CAVE-ETD runs in the Cave Automated Virtual Environment (CAVE), a 10x10x10-foot room, with stereoscopic projections on three walls and the floor, wherein the user may interact with the world through tracking devices, eyeglasses, and a wand. In CAVE-ETD, "books" are organized on "shelves," shelves are laid out in "aisles" in a "room," and rooms are labeled and arranged in a logical sequence. Books can be browsed on the shelves by navigating through the room and reading the titles on the book spines. Real-time clustering methods are being investigated to determine their utility in the CAVE. Both the CAVE-ETD and the 3DL rely on a user's prior knowledge and experience in a traditional library.

### D.3 ETD-DL Maintainers

**Approval Wizard.** Complementing the student-based ETD submission wizard, the Graduate School has similar software that helps them receive, annotate, and approve ETDs. Included with this package is the ability to update metadata fields (e.g., for correcting spelling errors and applying controlled vocabularies) and send email messages to students whose ETD chapters need to be updated. Also included is the ability of the Graduate School to revise availability restrictions on ETDs as students request such changes.

**Metadata.** Digital resources place new demands on library catalogers. Document metadata that used to consist of title, author, and subject, now must consider document type, size, submission date, and what applications are necessary for viewing. Keeping with new metadata standards, we have aligned the metadata fields for ETDs with the Dublin Core (<http://purl.oclc.org/dc/>). In the future we hope to leverage our investment in Dublin Core to share library metadata between NDLTD sites and with other online scholarly document initiatives. The NDLTD is an active member of the Open Archives Initiative (<http://vole.lanl.gov/ups/>).

**Handles as persistent URLs.** Having persistent digital object identifiers means that links to ETDs will survive machine changes, re-hosting of collections, and enable transparent collection mirroring. Under the

handle system (<http://www.handle.net/>), we have reserved major naming domain "ndltd" and have implemented "ndltd.vt" for Virginia Tech ETDs. Under this auspice, we assign a serial identifier to each ETD's "home page" in the Virginia Tech collection. Once handles are assigned and handle servers are in place, then ETDs can migrate from server to server without having to update published information.

## D.4 Community of Interest

In the grant period we have grown a large community of interest around ETDs, the NDLTD, and issues in electronic, academic publishing. Indeed one of our objectives was to spawn this community. "Joining the NDLTD" is a significant step for institutions to take that shows that they are interested in accepting ETDs and willing to address issues in technology, publishing, and the future of the academic promotion and tenure process.

We learned early in the process that universities act very slowly, even if it means wasting tens of thousands of dollars per year. The majority of our efforts went to making faculty and graduate schools aware of the process and showing them how they can save money.

Leading this community is the NDLTD Steering Committee (SC). The NDLTD-SC convenes twice per year, discussing issues in project evaluation, accomplishments, continuing growth, and is collectively conscious of new developments in the electronic, scholarly publishing industry. The NDLTD-SC plans alliances and recommends avenues for ongoing funding. Membership includes representatives from our member institutions as well as those with industrial interest, including Adobe Systems, IBM, and UMI. Information about participants and meeting minutes are available online (<http://scholar.lib.vt.edu/theses/NDLTD/>).

Related to the NDLTD-SC is the NDLTD Technical Committee. This team met for the first two years of the project, concentrating on technological issues related to NDLTD, ETD submission, ETD preservation, and software products and projects related to the process. In Summer 1999 it was decided to expand the committee structure of NDLTD and to replace this committee with several having more focused goals.

**Conferences and Symposia.** Several conferences and symposia have convened as part of or related to the NDLTD.

- ETD/NDLTD Workshop, MECCA Conference, Memphis, Tennessee, June 10-11, 1998 (<http://www.mecca.org/mecca-itec-conf/conftopics.html>).
- Virginia Tech hosted a series of seminars on electronic publishing in academe, Spring 1998 (<http://www.rgs.vt.edu/resmag/seminars.html>).
- The ETD User Group held a two-day meeting in May 1999 (<http://www.ndltd.org/ndltd-ug/990516.htm>).
- Building on the success of the ETD User Group 1999 meeting, the first annual ETD conference will be held in Tampa, Florida in March 2000 (<http://etd.eng.usf.edu/Conference/>). This event will include peer-reviewed papers about ETDs. Publicity efforts have included thousands of invitations.

**Visitors/Visits.** On almost 200 occasions, we have traveled to, met with, spoken to, and hosted interested institutions and potential partners—many more than we had expected or for which we had budgeted (Table 1). Visitors to Virginia Tech all paid their own way, and in some cases universities inviting the Virginia Tech team to speak paid some or all of the travel expenses.

**Advertisements.** We are competing for the attention of persons who are primarily disposed to other issues besides ETD submission. Despite unsolicited appearances in the Chronicle of Higher Education, New York Times (and other papers served by the API wire), National Public Radio, and numerous conferences with national and international attendance, our "product" recognition is still less than the

universal coverage we desire. While word of mouth has been most effective for us to relay product information, we have prepared supporting pamphlets and handed them out at scholarly gatherings and mailed them to graduate schools and libraries in universities nationwide.

**Table 1. Coverage by visitation, hosting, and membership, including 32 of 52 USA locations and 18 foreign countries and the United Nations.**

#### 32 of 52 USA Locations

**AL**, AK, AR, **AZ**, CA, CO, CT, **DE**, FL, GA,  
**HI**, IA, ID, **IL**, IN, KS, **KY**, LA, **MA**, MD,  
**ME**, MI, MN, MO, MS, MT, **NC**, ND, NE, NV,  
 NH, **NJ**, NM, NY, **OH**, OK, OR, **PA**, RI, **SC**,  
 SD, **TN**, TX, UT, VA, VT, WA, **WI**, WV, WY,  
**DC**, PR

#### 18 Countries and the United Nations

Australia, Canada, China, Croatia, Germany, India, Italy, Japan, South Korea, Mexico, Portugal, Russia, Singapore, South Africa, Spain, Taiwan, United Kingdom, UNESCO.

**ETD Sourcebook.** Working with editors from some of our member institutions, we have solicited more than ten chapters for the *ETD Sourcebook*. It will contain information about social and political aspects of ETDs, offering perspectives from graduate schools and libraries, library issues regarding collection management, and technical discussions covering PDF and ETD-ML. We plan to include portions of this FIPSE Project Report document as a chapter in the *ETD Sourcebook*, which we will publish in 2000.

## D.5 Membership Reports

The following three subsections are representative statements included in this report to illustrate activities at active NDLT D sites. Staff from those projects submitted the following statements.

**North Carolina State University.** Since Spring 1997, the NC State Libraries, the Graduate School and the Information Technology division have collaborated on a pilot project that allows selected theses and dissertations to be submitted, approved, and accessed electronically. The local effort is part of a larger project sponsored by the Southeastern Universities Research Association (SURA) and prototyped by Virginia Polytechnic Institute and State University (Virginia Tech) to create a digital thesis and dissertation repository.

Using software developed by Virginia Tech, students submit their works to the Graduate School as Portable Document Format (PDF) files. Once reviewed and approved, they are transmitted to the Libraries for cataloging and then accessed via the Libraries' Web server. Submission electronically is entirely voluntary.

*The NCSU report was submitted by Rick Liston, Director of Information Systems, The Graduate School, North Carolina State University.*

**West Virginia University.** West Virginia University began a formal investigation of the electronic distribution of master's theses and doctoral dissertations on September 11, 1997. A planning/implementation team was charged with the development of an ETD program at WVU.

Virginia Tech's successful ETD and NDLT D programs were reviewed and the implementation team visited the Virginia Tech campus to discuss operational program issues with their staff. With the background gained from VT's program, The WVU implementation team formed three working groups to address major program development issues in the areas of technology, education, and policy/procedure. The result was the development of a program, modeled after the Virginia Tech ETD program, which was designed to fit into our local campus environment. On August 15, 1998, West Virginia University

implemented a policy that all masters' theses and doctoral dissertations must be submitted and archived electronically.

*The WVU report was submitted by Ruth Nellis, Coordinator of Library Renovation and Building Projects, Office of Academic Affairs and Research, West Virginia University.*

**Massachusetts Institute of Technology.** MIT began to build an online thesis collection in 1997, borrowing software and infrastructure from the Networked Computer Science Technical Reports Library (NCSTRL) project. The Web site went public in the autumn of 1998 (<http://thesis.mit.edu/>). Six-hundred dpi bi-tonal TIFF images, scanned from archival microforms to facilitate the delivery of printed copies on demand, are "harvested" for display (GIF) on the E-Thesis Web site, as well as subsequent print or e-delivery (PDF) requests. In late 1998 and 1999, a pilot project tested the feasibility of electronic submission of new theses at MIT, yielding fewer than twenty titles. MIT is in the process of becoming a member of the NDLTD.

*The MIT report was submitted by Keith Glavash, Head of Document Services, MIT Libraries.*

## D.6 Evaluation Design

Devising good ways to measure our own success (and digital library success in general) has become a significant part of the work of the NDLTD. Of particular note is that for DLs, it is difficult to measure how patrons are using the materials offered—due primarily to the ethical necessity of providing anonymous and unfettered access to a DL of this nature.

What should the evaluation show? With our evaluation we wanted to show that graduate education is being enhanced, that there is significant mass of transferred technology, and that time and money are being saved by students, libraries, and graduate schools. We show this in the NDLTD by demonstrating that graduate students and faculty are learning about contributing to and making use of the NDLTD, that there is extensive use of the NDLTD by others, and that time and money of students and universities are saved once an ETD infrastructure is established.

What instruments and metrics are involved? Our own evaluation instruments include logging accesses, collecting surveys, holding focus groups, and doing usability studies of the library services. We also have incorporated action research methodology to good effect. Details of each of these instruments and the metrics they use are discussed below. Results of the analyses are shown in Section E. We group the design into two parts, one for NDLTD in general and one specifically for ETDs at Virginia Tech.

### *NDLTD in General.*

1. **Action research.** Project workers are asked to reflect on the project as they see it—from the trenches of its ongoing development. This qualitative method measures how well we support the needs of ETD authors, institutions, and researchers.
2. **Federated search logs.** We analyzed log data since the inception of the federated search system to show how the service is being used.
3. **Focus groups.** Regarding annotation of ETDs and other ETD-related tools, Todd Miller held focus groups and conducted individual interviews collecting requirements and proposing solutions for increasing the usefulness of NDLTD.
4. **Membership growth.** The NDLTD is concerned with growing its membership. While joining is "free," it does require a thoughtful commitment on the part of the joining institution. Measuring the pattern of joining shows how well we are doing with dissemination of project ideals to ensure open access to ETDs by all. Noting patterns in type or location of joining institution should also contribute to our better growing our membership.

5. **Member institution survey.** Periodically, we have surveyed our member institutions, asking about their status, progress, and what concerns they have that are still outstanding.
6. **Usability study.** Several human-computer interaction researchers did a usability comparison of NDLTD, ACM-DL, IEEE-DL, and NCSTRL.
7. **Web site logs.** We also record accesses to our training, policy, and other supplemental materials. This provides additional information about progress of the project.

### *Virginia Tech ETD Collection*

1. **Approval form survey.** At the time of submission of an ETD at Virginia Tech, along with the approval form signed by student authors and their faculty committee, we ask five survey questions:
  - a. Have you or will you publish article(s) in peer-reviewed journals based on your ETD?
  - b. Do you plan to publish a book or manuscript based on your ETD?
  - c. If "yes," when do you expect to submit the manuscript(s) for publication?
  - d. If you chose not to release your entire work immediately for worldwide access, what caused you to choose that option?
  - e. Would you have been willing to release parts (i.e., chapters) of your ETD worldwide if access to the other parts of the ETD could remain limited?
2. **Conventional access statistics.** To show that ETDs were totally dominant over paper copies, we analyzed data collected from Virginia Tech library circulation and interlibrary loan of theses and dissertations.
3. **ETD collection logs.** We record the IP address and hostname of each access to the Virginia Tech ETD collection, whether the abstract is examined, and whether all or part of the full document is retrieved. We categorize these IP addresses by country of origin, day and time of access, and number of bytes downloaded. Comparing accesses year by year gives a comparison of how interest in our collection is changing.
4. **ETD patron survey.** Our original plan required that all library users identify themselves so that we could keep detailed usage data. However, because mandatory user identification is contrary to expected and ethical conduct in libraries, we have made the Virginia Tech ETD collection available without mandating that users supply identity and demographic information. Consequently, our electronic library usage is very high, robot crawlers can index ETDs, and our servers do not bear the burden of running a CGI program with each server interaction. Our modified evaluation plan includes an optional survey that collects demographics of ETD users and asks if they are using the materials for a course, self-study, research, distance learning, or other. In addition, we provide the facility for open feedback: the project team's and Web administrators' email addresses are available on the Web site.
5. **ETD student pages Web logs.** We record accesses to the ETD student site (<http://etd.vt.edu/>). Comparing accesses tells what part of the site is used and by whom.
6. **Graduate School time and cost analysis.** We interviewed Virginia Tech Graduate School staff for their first-hand report of savings of ETDs versus paper.
7. **Post-graduate survey.** The Virginia Tech Graduate School surveys some of its graduates to determine if they have derivative publications from their ETDs.
8. **Student time and cost analysis.** We estimated costs for the average paper thesis and dissertation and compared them to this year's typical ETD development scenario.

9. **Workshop attendance and survey.** We hold two workshops per semester (Fall, Spring, Summer) at Virginia Tech to help train students, faculty, and staff about the ETD process. We monitor the attendance and questions that the attendees ask.

## **D.7 Future**

**Society.** Incorporation of the NDLTD as non-profit, non-governmental organization is planned for January 2000. Incorporation should enable additional UNESCO support (UNESCO joined NDLTD early in 1999) and may lead to ongoing funding from such organizations.

**Technology.** We hope to improve and distribute browsing interfaces to the collection by collecting more user suggestions and incorporating those into the collection management scripts. We hope to provide better federated-search access to the collection and to share indexes and metadata between installed sites. Collection and metadata mirrors also are being considered for preservation's sake. We are working to provide machine translation of documents into Spanish, Japanese, and to and from other languages to help ensure technology transfer worldwide.

**Evaluation.** Worldwide we will continue to monitor the ETD collections and continue to collect and reply to email commentaries. At Virginia Tech we will continue to survey graduating students and request post-graduate responses.

**Dissemination.** NDLTD is committed to continued growth, supporting more and more universities with ETD submission software and helping them plan adoption of the ETD submission and maintenance process. We do this by keeping our Web site updated and making our submission and approval software more robust. We also plan to collect software from other NDLTD members to promote wider sharing.

We hope to maintain our positive momentum by securing continued funding from FIPSE and other sources in the coming months.

## ***E. Evaluation and Project Results***

From the following, we argue that our project efforts have been successful. Thus: (1) due to our efforts students and faculty are learning about electronic publishing, (2) technology is being transferred to others, and (3) students and institutions are saving time and money.

### ***E.1 NDLTD in General.***

#### **Action Research**

The following assertions emerged over the course of the project.

**Widespread adoption.** For ETDs to be accepted everywhere, supporting technologies need to be integrated and use encouraged at every work point. As Rick Liston (NCSU) reports, the ETD project is a success waiting to happen when, and if, submission is mandated. Virginia Tech experience supports this as well. Students were unhappy with the mandatory submission at first (initial change is challenging), but now they are beginning to request electronic submission of homework assignments and final papers as part of their regular classroom activities.

**Dissemination.** One or more face-to-face meetings are required between our project team and another site to ensure joining and continued support of the NDLTD.

**Community.** The NDLTD process has been analyzed as part of researcher Kipp's dissertation (May, 2000). Kipp characterizes the NDLTD as a "community" type digital library, one that supports the ongoing activities of a community, in particular serving as its community memory. Serving the ETD community involves these activities:

- training authors in document development and warehousing their works,
- addressing broad social issues related to the purpose of the community,
- supporting communication within the community, and
- perpetuating itself.

**XML.** Within the NDLTD, markup language adoption is still alarmingly weak. It is not that XML tools are not available to those interested in new technologies. Rather, it is because XML is not the most common document preparation environment. It will be very difficult making XML submissions mandatory until the most popular word-processing tools support XML.

**Authoring.** Though improving, many students' word processing skills are not on par with successful thesis development. This is an opportunity for further research in graduate education.

## Federated search logs

For the period April 13, 1999 to November 30, 1999, there were 7420 federated searches of the NDLTD sites, including requests for "polymer photovoltaic cells," "factors related to computer use by teachers in classroom instruction," and "telecommunications in China."

## Focus groups

Digital library tools designers (students and faculty in Computer Science) came together to discuss NDLTD-related use cases and proposed how these new tools would behave. The focus groups worked on possible digital library tools (proposed earlier by the NDLTD project team), including: bibliography builder, ETD compare tool, open-problem finder, influential article finder, and annotating tool. The ETD compare tool and the influential article finder were most interesting to the focus group (however none surfaced as most useful), and an "idea-tracker" tool was proposed that would let readers of ETDs track ideas across an entire DL corpus (Miller, 1999).

## Growth of membership

Seventy members of the NDLTD are listed in Table 2. Of these, half are USA universities, more than one-third are international universities and less than one-sixth are interested institutions. Members come from all over the United States and all over the world. Several others are in the process of joining.

Membership in the NDLTD has grown smoothly since the first member joined in 1996 (**Figure 1**), but how will the trend continue? We cannot correlate spikes in joining with dates of flier mailings. Indeed, direct response to the fliers has been low (less than 1%). Nevertheless, we feel that continued production of fliers is necessary to support our verbal message and to provide a tangible reminder of the NDLTD after our presentations and visits. Clearly site visits can be correlated to growth in membership. Concentrations occurring in Summer of 97 and 98 can be attributed to visits, talks, and meetings. Growth in 1999 might be attributable to the cumulative effects of all ETD-related activities and efforts.

## Membership Survey

The following subsections report on a representative set of sites working on ETDs.

**WVU.** By the end of the first year of West Virginia Universities' program 257 ETDs had been archived. In the semester ending in August 1998 just 10 of the 86 submissions were in electronic format. However, only one of the submissions received in Spring 1999 was in paper format. At least 393 ETDs are on file in the Universities' ETD Archives. Master's theses comprise 57% and doctoral dissertations account for the remaining 43%.

**Table 2. NDLTD Membership (70).**

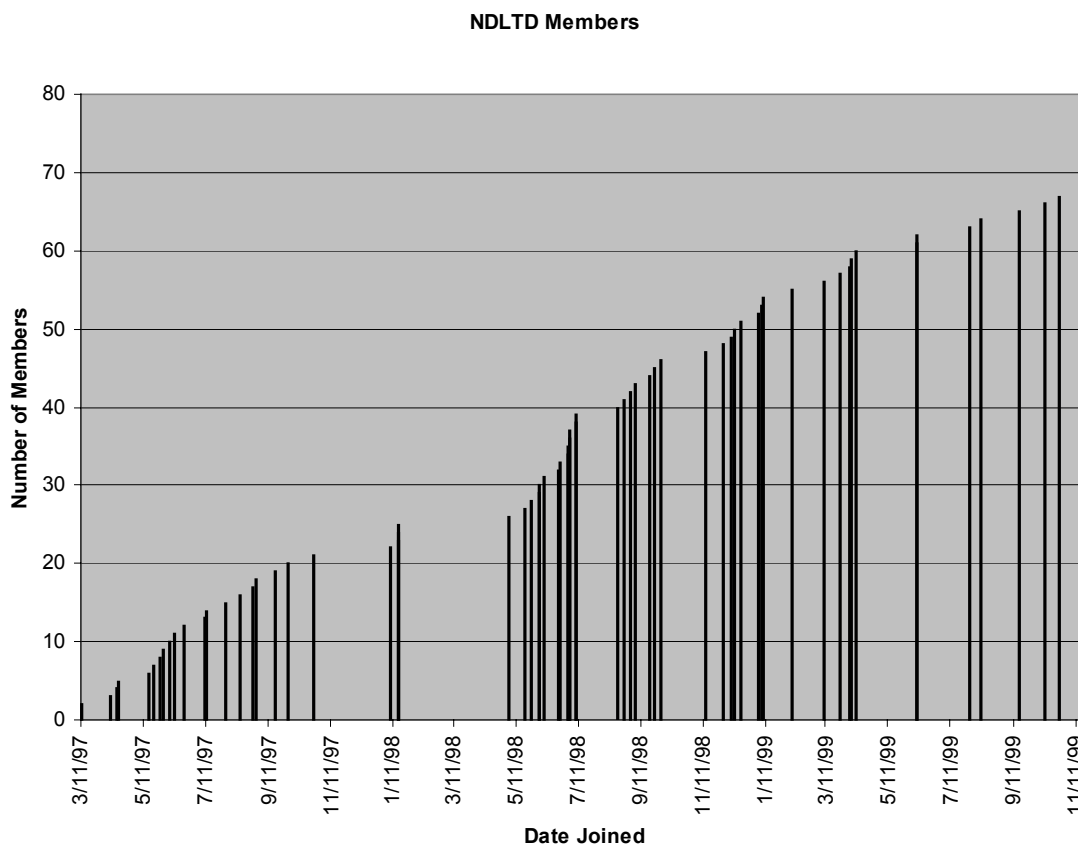
<b>USA Universities (35)</b>	<b>International Universities (27)</b>	<b>Institutions (8)</b>
Air University (Maxwell AFB, Alabama)	Chinese University of Hong Kong (China)	Coalition for Networked Information
Brigham Young University	Chungnam National U., Dept of CS (Korea)	Committee on Institutional Cooperation
California Institute of Technology	City University, London (UK)	Diplomatica.com
Clemson University	Curtin University of Technology (Australia)	Dissertation.com
College of William and Mary	Darmstadt University of Technology (Germany)	Dissertationen Online
Concordia University (Illinois)	Freie Universitat Berlin (Germany)	Ibero-American Sci. & Tech. Ed. Consort. (ISTEC)
East Carolina University	Griffith University (Australia)	Organization of American States (SEDI/OAS)
East Tennessee State University	Gyeongsang National University, Chinju (Korea)	UNESCO
Florida Institute of Technology	Humboldt-Universitat zu Berlin (Germany)	
Florida International University	Indian Institute of Technology, Bombay (India)	
George Washington University	Nanyang Technological U. (Singapore)	
Marshall University (West Virginia)	National Library of Portugal	
Miami University of Ohio (Oxford, Ohio)	National U. of Singapore School of Computing	
Michigan Tech	Rhodes University (South Africa)	
Naval Postgraduate School (Monterey CA)	St. Petersburg State Technical U. (Russia)	
North Carolina State University	Universidad de las Américas Puebla (México)	
Pennsylvania State University	Universidad Politecnica de Valencia (Spain)	
Rochester Institute of Technology	Université Laval (Canada)	
University of Colorado Health Sciences Center	University of Alicante (Spain)	
University of Florida	University of Guelph (Canada)	
University of Georgia	University of Melbourne (Australia)	
University of Hawaii, Manoa	University of New South Wales (Australia)	
University of Iowa	University of Pisa (Italy)	
University of Maine	University of Queensland (Australia)	
University of Oklahoma	University of Sydney (Australia)	
University of South Florida	University of Waterloo (Canada)	
University of Tennessee, Knoxville	Wilfrid Laurier University (Canada)	
University of Tennessee, Memphis		
University of Texas at Austin		
University of Virginia		
University of Wisconsin, Madison		
Vanderbilt University		
Virginia Tech		
West Virginia University		
Worcester Polytechnic Institute		

The impact of the ETD program on the West Virginia University campus has been far-reaching and diverse. It has ranged from the direct impact of theses and dissertations being accessible throughout the world to enhancing research collaboration. Additional grant proposals relating to the project have been developed. There is a heightened awareness of electronic publishing on campus. The effort has fostered an awareness of how electronic publishing impacts the disabled user and how we can improve the transfer process, providing the disabled user with easier access to electronic materials. Also we have seen improved campus coordination between the academic units and supporting campus service units.

The most far-reaching impact may lie in the Universities' new focus upon the campus' electronic information environment. Several university segments are beginning to address the total campus information environment and the benefits of an integrated approach to handling information.

**MIT.** The combined online collection, consisting of both the scanned and submitted e-theses, numbers over 3500 titles to date. The average usage of the site is about 6000 requests per day, 4000 of which are individual page images. The online collection currently expands at a rate of ten to twenty titles per week. The popularity of the site among both MIT and outside users continues to grow, raising the expectation

that all MIT theses will soon be available digitally. While that is unrealistic, it is prompting us to move ahead on two specific fronts: (1) scanning a subset of new theses from paper in anticipation of user interest, rather than waiting for a request to trigger the conversion to digital; (2) adding URLs to catalog records. In addition, we plan to continue to accept e-submissions on a passive basis, hoping to institute an active program when we have solved more of the political and technical challenges of electronic submission at MIT. It is our goal eventually to receive all theses in digital formats.



**Figure 1. Membership growth of the NDLTD**

NCSU. Submission electronically is entirely voluntary. Seventy-two (72) theses or dissertations have been submitted, approved by the Graduate School and publicly archived by the Library since project inception. One (1) is currently awaiting approval. Six (6) have been through the submission and approval process and are currently archived in a proprietary status. Excluding Internet robots/spiders as well as library and graduate school staff, the public ETDs have been accessed approximately 58,614 times since March 1997. The vast majority of these "hits" are the ETD's abstract files; the PDF versions of the documents were accessed each approximately 348 times.

The Graduate School Thesis Editor and Webmaster continuously address student issues relating to ETDs. The attitude among students towards electronic submission has changed from slightly negative to positive overall and the number of submissions has been on the rise as students finish their graduate programs.

**Dissertation.com.** Dissertation.com is part of Amazon.com and functions as a publishing agent for students. It offers electronic dissertations in PDF or paper formats for \$20 to \$40. Abstracts are freely available.

**University of Tennessee, Memphis.** The University of Tennessee, Memphis has three documents in its collection. Of these, all are in PDF, but one is also in HTML.

**University of Michigan.** While not an official member of NDLTD, the University of Michigan has begun a thesis pilot program. Instead of PDF, they have four ETDs in SGML, conforming to the Text Encoding Initiative Document Type Definition ("TEI," Sperberg-McQueen and Burnard, 1994).

**University of Virginia.** The University of Virginia has adopted an ETD pilot; it accepts electronic theses from Engineering bachelor's students. The university plans to require Master's theses and PhD dissertations at a later time. The current collection is stored in NCSTRL.

**International.** NDLTD accepts and encourages international membership. Approximately one-third of NDLTD members are not based in this country.

UNESCO is interested in electronic dissemination of theses and dissertations, convened a workshop on this in September 1999, and plans ongoing activities (see <http://www.unesco.org/webworld/etd/>).

Seven institutions in Australia, led by the University of New South Wales, and centered in its library, are collaborating to begin accepting electronic theses from postgraduate students. This "Australian Digital Theses Project" has standardized on SGML and PDF as document formats. The collection's oldest work is dated 1968.

The project "Theses Online," sponsored by the German Research Foundation (DFG) and initiated by a subgroup within the Initiative of the German Learned Societies for the Advancement of Digital Information and Collaboration, started in spring 1998 and was terminated in March with a conference held in Jena, Germany. Funds for a second project were granted by the DFG with a heavy emphasis on the collaboration with libraries and university computing centers, with research and development running from March 1999 to March 2000 at a volume of EU 300,000. Among the learned societies involved in the project are chemistry, computer science, education, mathematics, and physics, involving five German universities. Participants in the second proposal are also computing centers, libraries and the German National Library (DDB). Please see [http://www.educat.hu-berlin.de/diss\\_online/englisch/index1e.html](http://www.educat.hu-berlin.de/diss_online/englisch/index1e.html).

Encyclopaedia Diplomica is a German company acting as a selling agent for students who prepare scholarly works. Papers are in one of the following formats: Word, PDF, or PostScript. Abstracts and full tables of contents are available for free. Prices for the full documents are the equivalent of 150 to 300 US dollars. The collection offers approximately 20 titles. Most of the documents are in German; the rest are in English or French.

The Rhodes University of South Africa has begun an ETD pilot project. They request both paper and digital submissions.

The University of Waterloo in Ontario, Canada coordinated a three-institution cooperative and has sixteen documents online, in PDF and PostScript sources, including one that is dated 1964. The Electronic Thesis Project Team and the University of Waterloo Library sponsor the local site. They provide documents for free, but request the name, affiliation, and "reason-why" from the patron before permitting the thesis to be downloaded. The site uses OpenText for searching the full text of the collection.

## Usability study

Digital library users (N=48) were asked to work with four digital libraries in a study undertaken Fall 1997: ACM, IEEE-CS, NCSTRL, and NDLTD. While the NDLTD had the fastest search time, the forced choice of "Browse" versus "Search" was found to be misleading. Other metrics included: breadth of coverage, deep historical content, full-text search capability, depth of coverage, readability of online text, timely content, and visual similarity to printed version (Kengeri, *et al.*, 1999). One notable result was that user expectations and priorities regarding digital library services varied dramatically over the several-hour

period in which they participated. Apparently students learn a good deal about the capabilities of a digital library during a short period of use.

## Web site logs

**www.ndltd.org.** Our project pages (<http://www.ndltd.org/>) are very popular (see Figure 2).

**www.theses.org.** Access counts to the digital library starting point (<http://www.theses.org/>) also are quite high (see Figure 3).

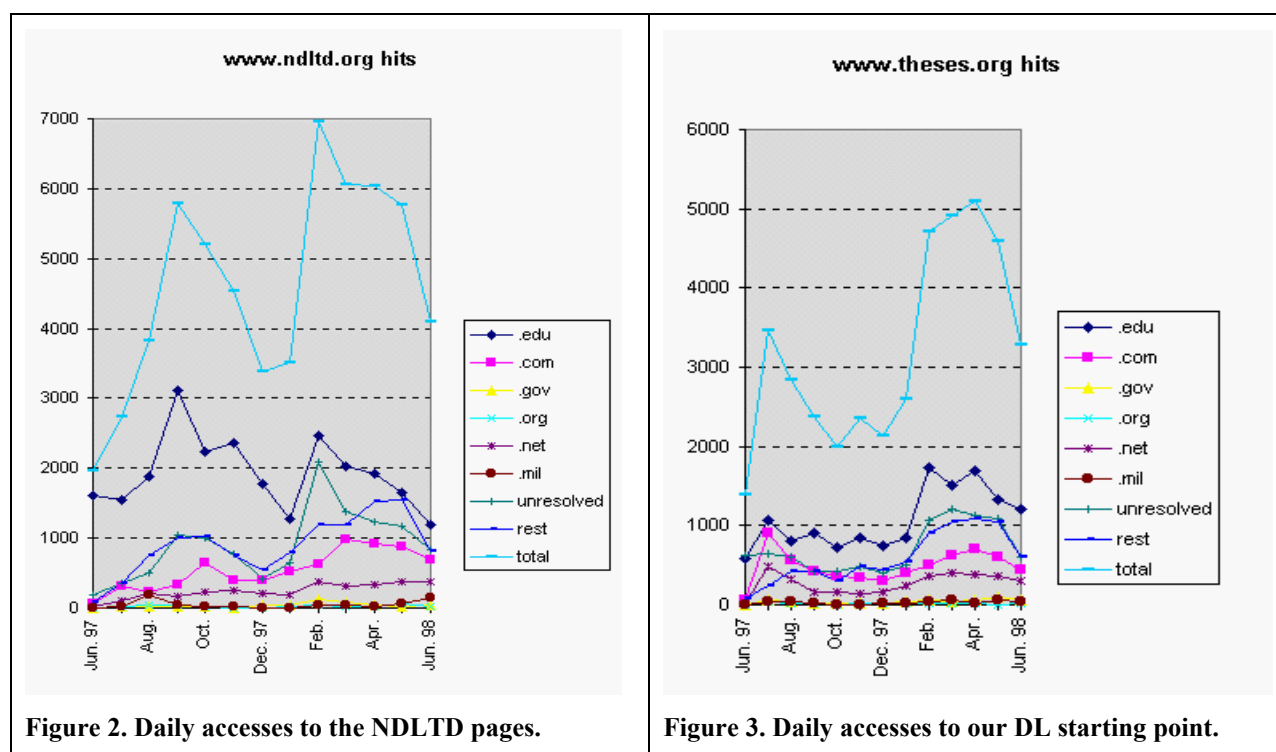


Figure 2. Daily accesses to the NDLTD pages.

Figure 3. Daily accesses to our DL starting point.

## E.2 Virginia Tech ETD Collection

### Approval form survey

On the approval form survey (tabulated spring 1999), authors:

- had NOT yet, but planned to publish articles in peer-reviewed journals (70%);
- did NOT plan to publish a book or were not sure they would (80%);
- if publishing, planned to publish within 12 months (2/3);
- if withholding worldwide access, were doing so because of planned publication of parts or whole (3/4); and
- were not willing or not sure that they would release specific chapters of their ETDs for worldwide access (90%).

### Conventional accesses

Virginia Tech's most frequently accessed dissertation had 75,340 log entries in 1998. Allowing one checkout per week, it would take hundreds of years and considerable expense to serve this many requests

in a conventional library. Alternatively, copying and shipping CD-ROM versions of the dissertation would cost tens of thousands of dollars.

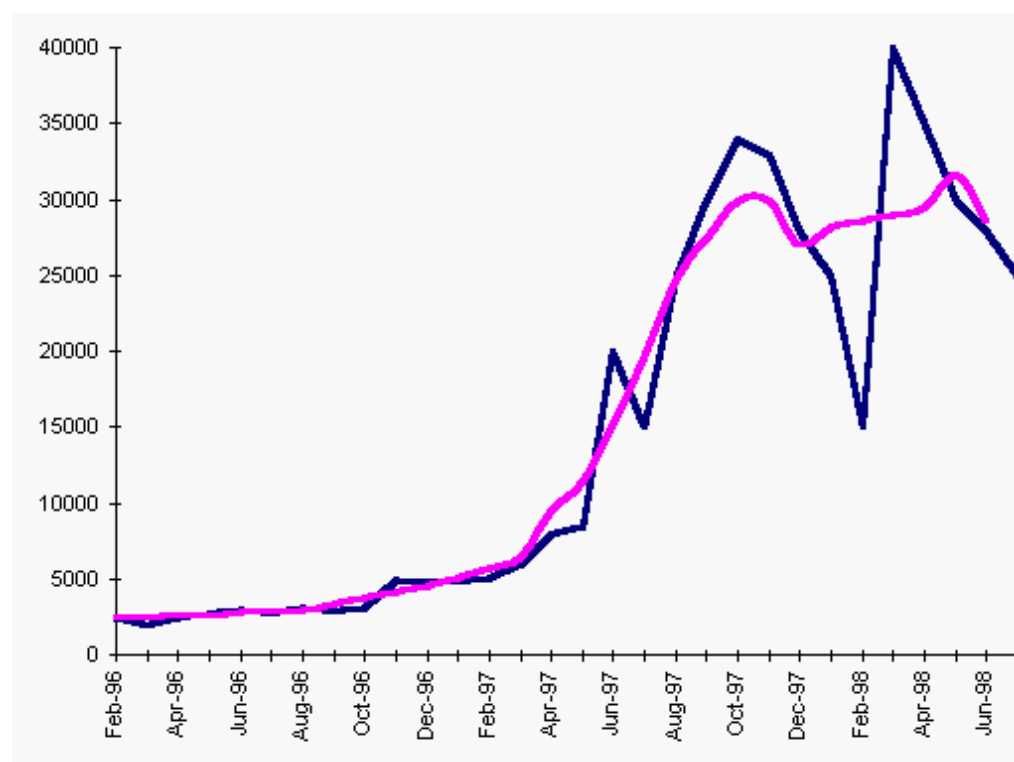
Based on current submission trends, Virginia Tech easily can accept six more years of ETDs (about 8 gigabytes) such that the additional cost per ETD is negligible. For an additional server cost of \$5000 we estimate that we can receive and serve ETDs through at least 2020. Also note that patrons cannot lose or damage the original library copy of ETDs.

**Table 3 ETDs at Virginia Tech circulate as much in one week as TDs did in five years.**

Document Type	Corpora	Total Circulations	Duration
Paper dissertations	4000 documents	8000	5 years
Paper theses	5000 documents	7000	5 years
ETDs	2000 documents	7000-8000	ONE WEEK

### ETD Collection Logs

Accesses to the VT-ETD collection have risen from 2500 per day to 30,000 per day, showing that these documents have amazing worldwide popularity (Figure 4).



**Figure 4. Daily accesses to the Virginia Tech collection.**

**ETD sizes overall.** ETD size ranges from 49 kilobytes to 80 megabytes, following an S-curve shown on a log scale in Figure 5. Because of the long tail, the mean sized ETD is bigger than one megabyte, yet the median ETD is only 800 kilobytes. Almost all ETDs, however, are between 100 kilobytes and 10 megabytes.

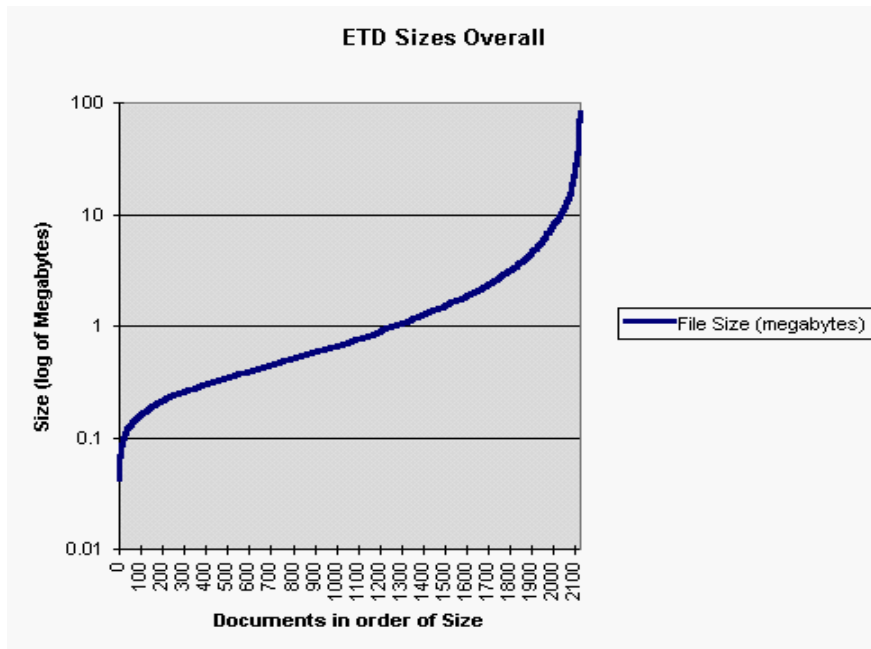


Figure 5. ETD Sizes Overall (log scale).

**Growth of ETDs.** Measuring submissions month by month, we found that the average ETD size is getting larger, while the median ETD size is staying about the same (Figure 6). We attribute this to the fact that more and more technically savvy students are submitting rich multimedia objects with their ETDs, though most students are turning in an electronic document that has text and (color) figures requiring less than a megabyte of storage.

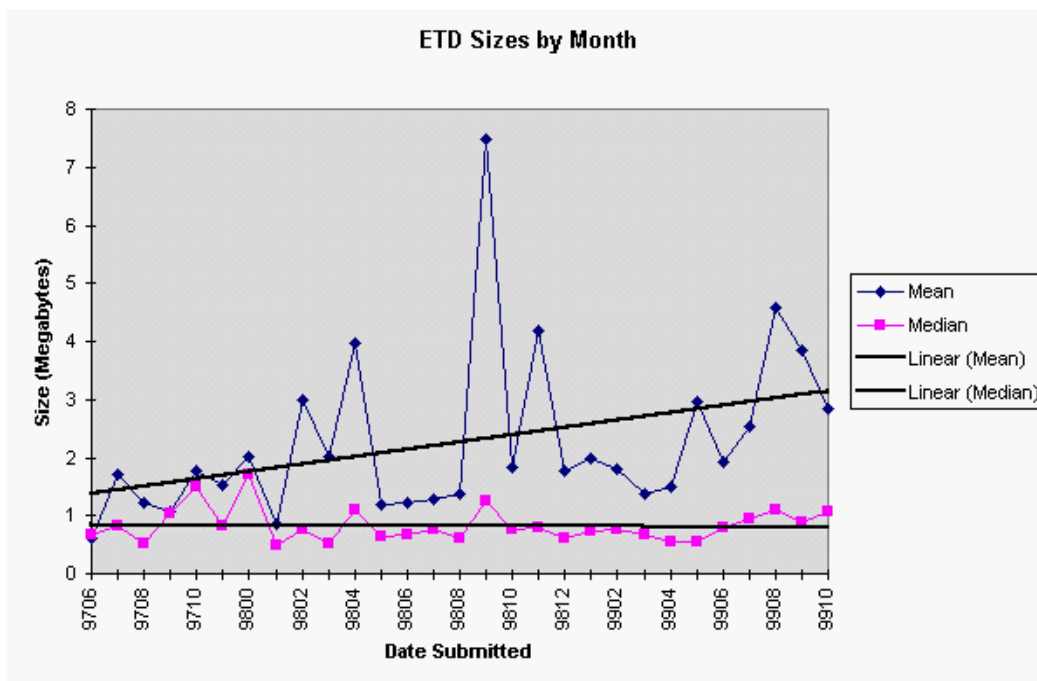
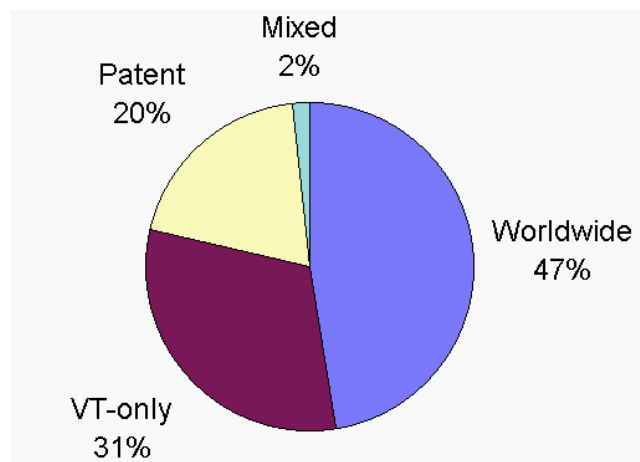


Figure 6. Sorted list of mean and median ETD sizes, in order by month of submission.

**Worldwide release.** A significant issue in ETD submission is whether to allow the electronic document to be viewed worldwide, by campus only, or not at all. The majority of students from Virginia Tech chose their document to be viewed by at least the campus (Figure 7). Ongoing discussions with publishers, so that they understand the goals and benefits of NDLTD, should help in this regard. See online discussion at <http://www.ndltd.org/publishrs/>. We are pleased to see a change in attitude by some publishers over the course of the project; the American Chemical Society developed a policy more favorable to NDLTD as a result of lengthy discussions.



**Figure 7. Student and committee choice for ETD availability from Virginia Tech.**

## ETD Patron Survey

In a small sample (N=30) taken last spring, users of the NDLTD and VT-ETD have these characteristics:

- have university affiliation (70%) as a researcher (47%)
- use PCs (80%) instead of Mac (17%) or UNIX (3%),
- are familiar with PDF (77%),
- have not themselves submitted an ETD (70%)
- use the library for research or personal interest (60%),
- have viewed five or fewer ETDs (70%), and
- plan to use the library monthly or less often (63%).

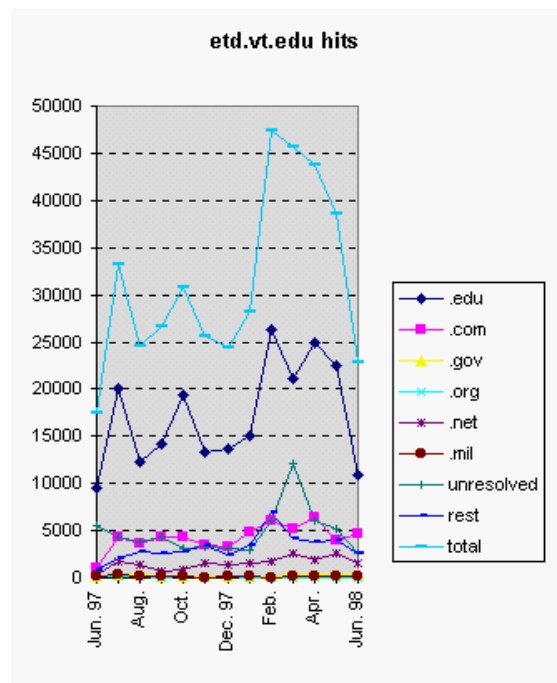
## ETD Student Pages Web Logs

**etd.vt.edu.** Site <http://etd.vt.edu/> is very popular with students and others across the Internet (**Figure 8**). In particular, our "How-to" pages (<http://etd.vt.edu/howto/>) for creating and submitting ETDs were the most popular within the ETD site (Figure 9). These pages have accounted for 20% of the accesses to this Web host, more than the submission instructions and the home page itself. Thus we conclude that our efforts to train students to be better electronic publishers have succeeded.

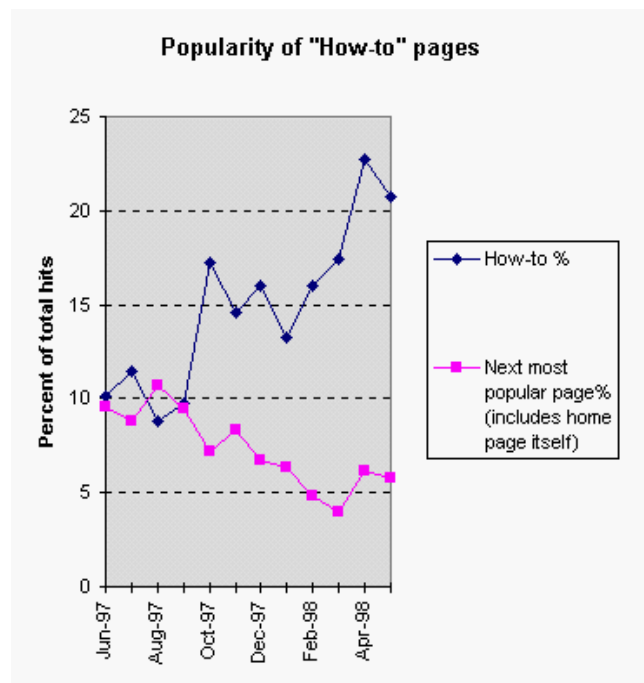
## Graduate school time and cost analysis

Gwen Ewing of the Graduate School approves the majority of Virginia Tech ETDs. She reports that she can review a one-file ETD in less than 20 minutes; ones with two more files requires upwards of 30 minutes, largely depending on use of hyperlinks, each of which she checks. One big saving is in turnaround time. ETDs with problems take days to resubmit; paper TDs with problems took 3-8 months to finally accept. Furthermore, some TDs with problems were sent to the library anyway, but ETDs with problems are rejected. The Plans and Clearances Office of the Graduate School is no longer crowded with

five shelves of double-stacked boxes of problem dissertations. While in the paper days, students could leave their three copies at the window during lunch in hopes of less careful scrutiny; now each ETD is reviewed and accepted with consistent criteria.



**Figure 8. Numbers of access to pages with information for students.**



**Figure 9. "How-to pages" compared with rest of accesses.**

## Post-graduate survey

Gwen Ewing of the Virginia Tech Graduate School reports that in October 1999, 410 letters with surveys were mailed to ETD authors with copies going to the committee chairs. Of the signed forms received back at the graduate school:

- 12 PhD authors requested release worldwide.
- 22 Master's thesis authors requested release worldwide.
- 77 requested to HOLD another year.
- 91 were returned with the student's signature, waiting for additional response from committee chair.
- The remaining authors are still undecided.

Because it is increasingly difficult over time to contact graduated students and their faculty committees, Virginia Tech plans to change access policies so that almost all works will become available automatically after some reasonable period of time, aimed to accommodate publisher priorities, say three years.

## Student time and cost analysis

Direct cost to the student can be significantly less than in the days of paper—even free—when the universities provide the PDF tools in their laboratories. Although student pricing (that began at \$45) for Adobe Acrobat has roughly doubled over the course of the project, students enjoy significant cost savings by not having to make up to five paper copies (on bond paper) for initial submission. Resubmission is where students can save the most, as changes to electronic files can be made online for no added cost.

## **Workshop attendance**

Workshop attendance at Virginia Tech has slowly declined over the past years. The largest was in the semester of the initial Virginia Tech requirement, with some 300 people attending. In each 1999 workshop, we had an average of 30 people. Furthermore, the questions are changing. Rage and indignation (from a small but vocal group reacting to Virginia Tech's being the first to require ETD submission, with only sketchy knowledge of the initiative) were expressed early in the process (1996-98), but after four semesters with some discontent (declining gradually), workshop participants now only bring very specific concerns about technology and PDF. Particularly, they are asking questions about getting specific application file formats into a standard accepted by the Graduate School to enhance preservation possibilities. They also are interested in getting help understanding the steps required in the graduation process.

## ***F. Summary and Conclusions***

**Summary.** NDLTD has made a significant and ongoing impact on higher education and the scholarly publishing process.

**Insights for others.** Others whose projects encourage membership and joining by institutions should know that universities generally adopt conservative, wait-and-see attitudes. Even though universities are wasting significant amounts of money by delaying the electronic thesis process, instantiating those processes and effecting an institutional change happens very slowly. Representatives require face-to-face contact to be "convinced" that the change will not be met with uncomfortable opposition. Representatives must load their own guns with arguments for every critique—after all, university faculty are in the critiquing business. Furthermore the attention and loyalty of key parties involved (i.e., graduate students and their faculty committees) are much stronger to their disciplines than to their departments or universities.

## ***Appendices***

1. Information for FIPSE
2. Selected Bibliography

## **Appendix 1. Information for FIPSE**

### **Assistance from FIPSE**

FIPSE provided highly effective assistance to this project in a number of ways. Most important was the close cooperation and assistance of the Program Officer. Our program officer served as a member of our Steering Committee, active through attendance at meetings of that group and through email and phone communications that related. The NDLTD Steering Committee, which met in D.C. twice a year throughout the project, gave high-level guidance for the overall effort. They reviewed our progress and plans, suggested the most important directions to follow, and helped set priorities. The need for backup materials for the meetings forced periodic review of accomplishments and remaining requirements, and the Steering Committee's comments on those provided very specific suggestions.

Second most important in terms of assistance by the Program Officer was communication during private sessions in connection with the fall FIPSE project directors meetings. Set in the context of work of other FIPSE projects, these comments helped provide overall guidance in terms of meeting objectives of the Department of Education in general and this project in particular.

Aside from the guidance of the Program Officer, FIPSE help at the annual project directors meetings was the area where FIPSE assistance was most valuable. These events are of high quality and not only provided a way for several of our co-PIs to learn more about related efforts, but also gave us a chance to contact people at other universities, to invite them to join NDLTD.

Finally, it is appropriate to state that the fact that FIPSE supported the project opened numerous doors to guarantee project success. Our initiative would not have been accepted by other universities and institutions, at home or abroad, without the "FIPSE seal of approval." The funding by FIPSE led to an order of magnitude more support, from corporations (especially IBM, Microsoft, Adobe, and OCLC) as well as others serving on the NDLTD Steering Committee and the members of NDLTD.

### **Future Proposals in this Area**

The US Department of Education has not been as active as it should be, either in the area of supporting graduate education, or in applying digital library technology. We found it exceedingly difficult to find other sponsors willing to sponsor any project focused on graduate education, no matter how meritorious. FIPSE should encourage and give priority to novel efforts aimed at this level of secondary education since there are almost no other opportunities for funding.

Regarding digital libraries, this is an important technology that has great value for supporting higher education. NSF has been very aggressive in this regard, through its Division of Undergraduate Education, and will provide support at the level of tens of millions of dollars, emphasizing science, mathematics, engineering, and technology. FIPSE should either work with NSF or develop its own program or sub-program in this area, focused on how digital libraries can support post-secondary education, especially in areas beyond the scope of NSF. A key person to work with in this regard at NSF is C. Dianne Martin.

Evaluation of projects in this area should make use of the technical experts in this field. People such as Steve Griffin at NSF have identified a pool of potential reviewers. Other aspects of selection of such projects could follow standard FIPSE methods.

### **Other Comments**

This type of project is exceedingly challenging, but highly rewarding. Integrating technology and education, encouraging collaboration among universities, and supporting universities to develop shared

infrastructure to each help their students move into the information age — all are of great import. There is a unique opportunity now to encourage universities and colleges to work together, to share resources, and to use technologies like digital libraries to facilitate this. We hope that FIPSE will contribute further to making this opportunity become reality.

## Appendix 2. Selected Bibliography

- Murat Bayraktar, Chang Zhang, Bharadwaja Vadapalli, Neill A. Kipp, Edward A. Fox. "A Web Art Gallery," *Proceedings Digital Library'98*, The Third ACM Conference on Digital Libraries, Pittsburgh, PA, June 23-26 1998, pp. 277-278.
- John L. Eaton, Edward A. Fox, Gail McMillan, "The Role of Electronic Theses and Dissertations in Graduate Education," *CGS Communicator*, January 1998, pp. 1,3-4, [http://www.cgsnet.org/Communicator\\_Online/Jan98/Lead198.html](http://www.cgsnet.org/Communicator_Online/Jan98/Lead198.html)
- John L. Eaton, Edward A. Fox, Gail McMillan, "Electronic Theses and Dissertations (ETDs) and Their Contribution to Graduate Education," *Proceedings of the 53rd Annual Meeting, mags*, Midwestern Association of Graduate Schools, 1997, pp. 74-79.
- Edward Fox, "Contribution by Edward A. Fox regarding Networked Digital Library of Theses and Dissertations (NDLTD) for *UNESCO Meeting*," September 27-28, 1999, Paris, France, <http://www.unesco.org/webworld/etd/contributions/fox.rtf>
- Edward Fox, "Networked Digital Library of Theses and Dissertations," *Nature, Web Matters*, August 12, 1999, <http://helix.nature.com/webmatters/library/library.html>
- Edward A. Fox. "The 5S Framework for Digital Libraries and Two Case Studies: NDLTD and CSTC," preprint for *Proceedings NIT99*, Taipei, Taiwan, Aug. 1999.
- Edward A. Fox. "Networked Digital Library of Theses and Dissertations," *Proceedings DLW15*, Nara, Japan, July 1999, <http://www.ndltd.org/pubs/dlw15.doc>
- Edward A. Fox. "From Theory to Practice in Digital Libraries: 5S and Educational Applications (NDLTD, CSTC)," *Proceedings of NSF-CONACyT Workshop*, Albuquerque, NM, July 1999, <http://www.ndltd.org/pubs/conacyt99.doc>
- Edward A. Fox, Gail McMillan and John L. Eaton. "The Evolving Genre of Electronic Theses and Dissertations", in *Proceedings Digital Documents Track of HICSS-32, Thirty-second Annual Hawaii International Conference on Systems Sciences (HICSS)*, Maui, HI - January 5-8, 1999, <http://scholar.lib.vt.edu/theses/presentations/Hawaii/ETDgenreALL.pdf>
- Edward A. Fox. "Networked Digital Library of Theses and Dissertations: A Framework for East-West Collaboration" *Proc. East Meets West, An Asian Digital Library Workshop*, 5-7 August 1998, Hong Kong, <http://www.ndltd.org/pubs/Asia98.htm>
- Edward A. Fox, Robert Hall, and Neill A. Kipp. "NDLTD: Preparing the Next Generation of Scholars for the Information Age," *The New Review of Information Networking*, 3: 59-76, 1997, <http://www.ndltd.org/pubs/nrin.pdf>
- Edward A. Fox, John L. Eaton, Gail McMillan, Neill A. Kipp, Paul Mather, Tim McGoingle, William Schweiker, and Brian Devane. "Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources," *D-Lib Magazine*, September 1997, <http://www.dlib.org/dlib/september97/theses/09fox.html>
- Edward A. Fox, Robert Hall, Neill A. Kipp, John L. Eaton, Gail McMillan, and Paul Mather, "NDLTD: Encouraging International Collaboration in the Academy," *DESIDOC Bulletin of Information Technology*, September 1997, <http://www.ndltd.org/pubs/dbit.pdf>
- Edward A. Fox, John L. Eaton, Gail McMillan, Neill A. Kipp, Laura Weiss, Emilio Arce, and Scott Guyer. "National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources," *D-Lib Magazine*, September 1996, <http://www.dlib.org/dlib/september96/theses/09fox.html>

Rekha Kengeri, Cheryl D. Seals, Hope D. Harley, Himabindu P. Reddy, Edward A. Fox. "Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD," *International Journal of Digital Libraries*, September 15, 1999, 2: 157-169.

Neill A. Kipp, "Case Study: Digital Libraries with a Spatial Metaphor," in *SGML/XML '97 Conference Proceedings*, Graphic Communications Association, December 1997.

Neill A. Kipp. "Document Type Definition for Electronic Theses and Dissertations," March 1997, <http://etd.vt.edu/etd-ml/dtdetds.htm>

Todd Hill Miller, "Improving Graduate Education Through Digital Library Tools," Bachelor's Thesis, May, 1999.

Constantinos Phanouriou, Neill Kipp, Ohm Sornil, Paul Mather, and Edward A. Fox, "A Digital Library for Authors: Recent Progress of the Networked Digital Library of Theses and Dissertations," in *Proceedings ACM Digital Libraries '99*, Berkeley, CA, Aug. 11-14, 1999.

James Powell and Edward A. Fox. "Multilingual Federated Searching Across Heterogeneous Collections," *D-Lib Magazine*, September 1998, <http://www.dlib.org/dlib/september98/powell/09powell.html>