

Submission for New Review of Information Networking - 1997

NDLTD: Preparing the Next Generation of Scholars for the Information Age

Edward A. Fox, Robert Hall, Neill Kipp
Virginia Tech, Blacksburg, VA 24016-0106 USA
fox@vt.edu

ABSTRACT

The NDLTD is an international consortium aimed at improving graduate education, enhancing universities' information infrastructure, and increasing scholarship by sharing through digital libraries. The NDLTD's success can be measured in different ways. For example, overall we can look at the number of universities that are members, and by the completion of milestones at each individual university. For the ETDs themselves, we can look at the number of ETDs received, and at the quality of those ETDs.

The NDLTD is emerging at a time when there are widespread efforts at establishing digital libraries. As digital libraries become central to the scholarly enterprise, it is necessary that all new works be prepared in electronic form.

The success of the NDLTD depends on adopting sensible strategies for the future. Standards must be developed, interoperability must be insured by overcoming both technical and social issues, and the NDLTD must demonstrate the necessary scalability to continue to grow.

During its first year the NDLTD has grown rapidly, opening its second year with over twenty members. There already has been encouraging interest in the ETD collection, and it appears that, on average, ETDs are accessed more than ten times as often as their paper counterparts. With continued corporate and academic support the NDLTD and its influence will continue to grow. We encourage everyone who reads this article to participate.

INITIATIVE

The Networked Digital Library of Theses and Dissertations (NDLTD) is an international consortium of universities and other interested parties aimed at improving graduate education, enhancing the information infrastructure in universities, and increasing the sharing of scholarship through digital libraries. It focuses on universities facilitating their students' efforts of preparing electronic dissertations, masters theses, bachelors theses, and related reports. If these students can learn about the publishing world, understand enough about electronic document preparation to prepare and submit their work to a

digital library, and make use of a digital library of theses and dissertations to extend their own research, the next generation of scholars will be much better prepared for the Information Age.

The first open discussion relating to the current initiative of preparing electronic theses and dissertations (ETDs) took place in fall 1987 when UMI invited representatives from ArborText, SoftQuad, University of Michigan, and Virginia Tech to Ann Arbor, Michigan. The group considered how the new standard, SGML (Standard Generalized Markup Language), and guidelines for applying it that had been developed by the Association of American Publishers, might relate to dissertations. Virginia Tech paid for development of a Document Type Definition (DTD) by Yuri Rubinski of SoftQuad that would be suitable for ETDs. Work continued slowly through 1992, when the Coalition for Networked Information, the Council of Graduate Schools, UMI, and Virginia Tech convened the first large workshop on this topic. Further support was provided through the Monticello Electronic Library, coordinated by the Southeastern Library Network (SOLINET) and the Southeastern Universities Research Association (SURA), that began a program for electronic theses, dissertations, and technical reports. SURA provided a grant to Virginia Tech for 1996-97 covering further SGML-related development and beta testing. The U.S. Department of Education gave funding covering 1997-99 for "Improving Graduate Education with a National Digital Library of Theses and Dissertations."¹ After the first year of that effort, with over twenty universities in NDLTD, including some in Africa, Australia, and Europe, NDLTD now refers to the Networked Digital Library of Theses and Dissertations, sometimes also called the ETD Initiative².

Goals

Many of the leaders in areas like education and research write a bachelors or masters thesis or a doctoral dissertation, and even more prepare some type of long report in connection with their period of involvement in higher education. Worldwide, the people involved in such efforts number well over 100,000 per year. Our **first goal** is to help with the education of these individuals, so they are prepared more effectively for the Information Age. We consider this an important responsibility of institutions of higher education, and so are striving to help those institutions accept and fulfill that responsibility. Thus, our **second goal** is to help universities develop an information infrastructure that will allow the first goal to be met. Our **third goal** follows from the other two, in that if hundreds of thousands of works are created electronically, and universities build up an adequate information infrastructure, the NDLTD will become a valuable aid to enhance sharing of information. It also should facilitate international collaboration³.

If the next generation of scholars is to become prepared for the Information Age, it is important to have a specific and feasible plan. At the heart of our approach is the observation that people learn best by doing. Thus, we believe that if a student can create his or her own ETD, add it to the distributed NDLTD collection, and access the growing

digital library of ETDs, then they will have achieved some base level of competency. In the process they will gain a number of key skills, and have increased their understanding of the Information Age.

We believe that the next generation of scholars should have at least a rudimentary understanding about the field of digital libraries, which brings together, in a form relevant to scholars, the many areas and sub-disciplines that relate to information (Table 1). These areas can be examined from a variety of perspectives. At the highest level is the view of degree programs (Table 1a); unfortunately in recent years there has been intense pressure in the U.S. to shut or downsize programs relating to library and information science. Just below degree programs are a small number of more specialized programs or minors (Table 1b). These programs are made up of directly related courses (Table 1c), and supporting courses (Table 1d) that could play a supporting role in understanding digital libraries. Finally, these courses include specific topics related to digital library research (Table 1e).

In the next subsection, we consider measures not only of the goal of preparing students for the Information Age, but also of other ND LTD goals.

<p>a) Traditional Degree Programs</p> <p>Computer Science Information Science Information Systems Library Science</p>	<p>d) Supporting</p> <table border="0"> <tr> <td>AI</td> <td>Geography</td> </tr> <tr> <td>Archiving and Preservation</td> <td>Image Processing</td> </tr> <tr> <td>Economics of Information</td> <td>Legal Issues of Information</td> </tr> <tr> <td>Cognitive Psychology</td> <td>OO Programming</td> </tr> <tr> <td>Communications</td> <td>Security</td> </tr> <tr> <td>Cataloging</td> <td>Sociology</td> </tr> <tr> <td>Education Techniques</td> <td></td> </tr> </table>		AI	Geography	Archiving and Preservation	Image Processing	Economics of Information	Legal Issues of Information	Cognitive Psychology	OO Programming	Communications	Security	Cataloging	Sociology	Education Techniques											
AI	Geography																									
Archiving and Preservation	Image Processing																									
Economics of Information	Legal Issues of Information																									
Cognitive Psychology	OO Programming																									
Communications	Security																									
Cataloging	Sociology																									
Education Techniques																										
<p>b) Specialized Programs/Minors</p> <p>HCI Multimedia Software Engineering</p>	<p>e) Topics</p> <table border="0"> <tr> <td>Agents</td> <td>IPR</td> </tr> <tr> <td>Browsing</td> <td>Knowledge Bases</td> </tr> <tr> <td>Classification</td> <td>Mediators</td> </tr> <tr> <td>Clustering</td> <td>Metadata</td> </tr> <tr> <td>Context Conversion</td> <td>Ontologies</td> </tr> <tr> <td>Digital Video</td> <td>OODB</td> </tr> <tr> <td>Document Modeling</td> <td>Privacy</td> </tr> <tr> <td>Electronic Publishing</td> <td>Routing</td> </tr> <tr> <td>Filtering</td> <td>Security</td> </tr> <tr> <td>GIS</td> <td>Storage</td> </tr> <tr> <td>Indexing</td> <td>Usability</td> </tr> <tr> <td>Interface Design</td> <td></td> </tr> </table>		Agents	IPR	Browsing	Knowledge Bases	Classification	Mediators	Clustering	Metadata	Context Conversion	Ontologies	Digital Video	OODB	Document Modeling	Privacy	Electronic Publishing	Routing	Filtering	Security	GIS	Storage	Indexing	Usability	Interface Design	
Agents	IPR																									
Browsing	Knowledge Bases																									
Classification	Mediators																									
Clustering	Metadata																									
Context Conversion	Ontologies																									
Digital Video	OODB																									
Document Modeling	Privacy																									
Electronic Publishing	Routing																									
Filtering	Security																									
GIS	Storage																									
Indexing	Usability																									
Interface Design																										
<p>c) Courses</p> <p>Database HCI Hypertext, WWW Information Retrieval Multimedia Networking Visualization</p>																										

Table 1: Areas of Study Relevant to Digital Libraries

Measuring our successes

Measuring success varies with activities. A student with simple requirements can learn enough to create and submit an ETD in less than an hour. For those with more complex requirements, such as when they feel obliged to use multimedia technology to convey the message of their research, a much longer investment in learning requisite concepts and skills may be required. Similarly, creating an ETD in Portable Document Format (PDF) is easier than creating one using SGML due to the scarcity of tools supporting SGML.

Thus, to measure the successes of the NDLTD it is important to break down the overall effort into a number of measures along a variety of dimensions. One dimension covers who or what is being measured: student, faculty, university, or ETD. Another dimension deals with how something can be measured: counting items, observing actions, analyzing surveys, testing skills, or gauging understanding. We briefly cover all these in a short narrative form below, to give the flavor of how the successes of the NDLTD can be assessed.

At the highest level, the NDLTD will be more successful if there are more universities participating. Our proposal to the U.S. Department of Education called for specific increases in the number of members in the U.S. at the end of each year of the three year effort. In addition, since international members have been joining too, we count the number of continents involved, and the number of countries. We also can measure the international spread of the NDLTD by looking at the number of languages in the collection, and the number of works in each language.

Regarding each university, one type of measure deals with milestones achieved. Are all of the following included in the effort: doctoral dissertations, masters theses, bachelors theses, and other student reports? Are submissions allowed, encouraged, or required? Is there a pilot; if so, how broad is its coverage? Is there a WWW site for information or an email address, phone number, or listserv for discussion? Are there training workshops for interested students and other events to train faculty? Is there an advisory committee representing campus groups to help plan and extend the initiative? Has an automated system been installed to simplify submission? Is there a procedure for backup and archiving of submitted works? Has a workflow process been implemented so that works flow smoothly from student through graduate school approval through cataloging through being made accessible (according to student and committee directions)? Is there a mechanism for assigning URNs (Uniform Resource Names) and for accessing available works through searching and browsing? Is access possible to the broad NDLTD through some type of federated search?

Regarding quantitative measures for universities, there are a variety of possibilities. What percentage of each of the types of works listed above is received electronically, out of the full yearly count? What percentage of the students has attended a training workshop? What percentage has attended an advanced training workshop (e.g., to learn about multimedia and SGML document preparation)? What percentage of the faculty has received what level of briefing or training? What percentage of the faculty has approved

an ETD and signed the appropriate approval form, understanding all the implications? What percentage of the students and of the faculty can browse, search, download and view an ETD - from their local collection, from the distributed collection, from works in PDF, from works in SGML, from works with various multimedia content?

Quantitative measures deal with the richness and variety of content and representation. Percentages are needed regarding numbers that: are in text-only form, are in PDF, are in SGML, use color, have a natural scene / image / photographic impression, have color graphics, include audio or video, have animation, include a simulation, contain a spreadsheet, provide raw data files, include internal hypertext links (e.g., table of contents, table of figures, table of tables, table of multimedia objects, glossary, references), include external hypertext links (using URLs or URNs), have Java or other program components, include VRML (Virtual Reality Markup Language) files, are dynamic (as opposed to traditional static documents), or include special PDF features (thumbnails, bookmarks).

Qualitative measures also are important. Some, from surveys, assess student reactions to the workshops, interfaces for accessing the NDLTD and the overall initiative. Surveys also convey an indication of the increase in skills and knowledge resulting from preparing an ETD. Of particular interest is ascertaining how comfortable students are with key concepts of document preparation (e.g., fonts, figures - bitmap vs. vector, hypertext, multimedia, compression, markup, capture, digitization, formats, format conversions, codes - ASCII or Unicode) and how much they understand the world of publication.

From NDLTD we hope to find data about scholarly work not yet available. For example, for each institution it will be helpful to know how many dissertations, masters theses, bachelors theses, and reports are prepared by students each year, in each field. It will be nice to know how many of these lead to one or more conference papers, journal articles, book chapters, or books - and how soon after (or before) the defense these are published. Regarding electronic accesses to ETDs, it will be helpful to count accesses to the equivalent of the cover page and abstract, as well as to the full work. That data can be sorted based on accesses from the local campus, as well as various categorizations based on time, geographical location, or logical sector (e.g., an educational institution vs. a commercial organization).

In addition to the basic data discussed above, we also must observe trends. We expect over time that more universities will join, that more students will contribute their works, and that the works supplied will get larger and have richer multimedia and hypertext content. We expect that the number of accesses to the NDLTD will continue to increase, and come from a more varied segment of scholars worldwide.

Given the above mentioned goals and measures, it is important in the next sections to explain the contextual situation for the NDLTD, discuss the strategy adopted to expand the NDLTD, review the progress made to-date, and summarize results.

CONTEXTUAL ISSUES

The NDLTD is emerging at a time when the Internet in general, and the WWW in particular, have caught the imagination of millions worldwide. In spite of the current excitement, the depth of understanding in these areas is woefully inadequate. In particular, few understand the underlying principles and technologies, and very few have training and skills that relate to library and information science. It is clear that we must give attention to these deficiencies. In addition, the NDLTD is being developed in the context of work on digital libraries, spread of tools for electronic document preparation, and upheaval in the publishing world, as discussed below.

Digital libraries

The field of digital libraries, while discussed openly in the 1960s⁴, has become well established only in the 1990s. ACM, the First Society in Computing, had its first international conference in this field in 1996 and will continue a large annual meeting into the foreseeable future. NSF, DARPA, and NASA funded a \$24 million Digital Library Initiative for the period 1994-98⁵, and NSF is coordinating a follow-on effort that will be roughly double the size, and have more of an international flavor⁶.

Worldwide digital library efforts number in the hundreds. Many are oriented toward research, but even more have a development flavor. Most deal with capture and conversion of some existing body of materials so that access can be controlled and extended. One example was the Computer Science Technical Reports (CS-TR) project⁷, that began with scanning all old technical reports at five large U.S. universities.

When the first version of the proposal to the U.S. Department of Education for the NDLTD was written in 1995, work on the Networked Computer Science Technical Report Library (NCSTRL-see <http://www.ncstrl.org/>), a merger of the former CS-TR and WATERS efforts, was well along⁸. It was clear (to author Fox, who had been involved since their inception in WATERS and NCSTRL) that a number of universities and other groups could collaborate to build a distributed digital library, and that technical solutions existed to make this feasible and useful. However, some improvements to the model seemed in order, and a more reliable mechanism for growth, management and quality assurance seemed necessary. In particular, consideration of electronic publishing, and the publishing world in general, is essential.

Electronic document preparation

For digital libraries to become established at the center of the scholarly enterprise, it is necessary that all new works be prepared in an archivable electronic form, instead of, or in addition to, any non-electronic forms that are appropriate. This is one of the reasons why the NDLTD is so important - it aims to ensure that those who will engage in

scholarly activities in the future will know how to undertake such electronic document preparation.

Almost every computer supports word processing, and almost every new document is prepared using some type of word processing or formatting software. Many public schools and public libraries have training programs on creating home pages or other documents for the WWW. Nevertheless, few new documents are entered into a digital library or archived electronically in a reliable fashion. Furthermore, while many people know how to create some type of electronic document, few can create a large, fully electronic document, suitable for archiving, and which applies hypertext and multimedia technologies optimally to convey the desired message, e.g., an ETD. Similarly, few people understand key concepts of document preparation (see “Measuring our Successes”, above).

One of the aims of the NDLTD is to prepare extensive online and written materials to help students learn what they need to know to prepare large, archivable, expressive, electronic documents. Using hypertext and multimedia methods, the materials at <http://etd.vt.edu/> provide an example of how students can learn key concepts and have complex questions and problems solved through suitable reference materials. In addition, tools are under development to make it easier for students to create documents in SGML, and discussions are being prepared to explain the key concepts of declarative markup⁹.

Publishing world

The NDLTD is emerging at a time of turmoil in the publishing world. Most traditional publishing, that relates to the reward system that young scholars face (e.g., tenure, promotion, and merit awards), takes place through paper journals or books. At the same time, a number of electronic journals have emerged, and many publishers are shifting to a dual-publishing model, whereby both paper and electronic documents are produced. In 1997, the more forward-thinking of those publishers have begun to apply digital library technologies to deliver their electronic holdings, e.g., American Association for the Advancement of Science, ACM, American Chemical Society, Elsevier, Institute of Electrical and Electronics Engineers Computer Society, and the National Academy of Sciences. This shift in production and delivery is new and expensive, increasing the costs of publishers in the short run, though in the long run, if electronic delivery becomes the norm, costs should decrease.

On the consumer side, libraries and individual readers also face benefits and problems as this shift occurs. Paying extra for electronic services is particularly difficult since for a decade there has been a yearly escalation in serials prices of almost 14%. Providing electronic services requires not only purchasing of subscriptions or services, but also involves developing a suitable delivery infrastructure (of computers and networks) and support staff (of technicians and specialists for planning, acquisition, and training). While universities are shouldering all these costs, many consider it wise to leverage their

investments by providing other services, like the NDLTD, which offer substantial benefits with very little additional expense.

Because the technologies and procedures are new, it is difficult for some publishers to develop new policies and for consumers to understand these. One of the aims of the NDLTD is to assist universities with educational activities aimed at clearing the air regarding such issues as copyright, right of first publication, handling of derivative works, and distribution of preprints or reprints. In particular, because UMI has published doctoral dissertations and sold copies on demand for nearly one hundred years, it is hoped that universities will be able to freely share most theses and dissertations without complex discussions with publishers over each of the hundreds of thousands of works of this type provided yearly.

STRATEGIC ISSUES

The success of the NDLTD depends on adopting a sensible strategy that involves standards, ensuring interoperability of participating systems, and implementation of mechanisms for scalability to encompass thousands of universities worldwide, hundreds of thousands of student authors, and millions of users.

Standards for sharing

Since 1987, the use of standards was considered essential for ETDs. Three classes are of particular importance. The first type of standard deals with electronic document preparation. At a 1994 meeting of interested universities, it was decided that ETDs would be represented both in PDF and SGML, as well as with relevant multimedia standards. The PDF version is easy to prepare, and allows a university to begin in the project quickly. It provides many benefits, and is preferred by those who wish a fully rendered version of the work to be archived, so no doubt will exist as to the author's intended presentation. SGML affords other advantages, including flexibility of delivery (as client platforms and preferences change), context-dependent searching, and higher certainty of archivability. SGML documents also provide a framework to link together a variety of multimedia works, such as those with sound and video, which do not fit into PDF files.

It is clear that multimedia standards such as JPEG for images, MPEG for video, and VRML for virtual reality files, can be archived. However, use of proprietary formats like those used with AuthorWare, Director, spreadsheet packages, etc., provides no guarantee that a file will be intelligible in a few years, and even less certainty of utility in 10, 50, or 200 years. Thus, using SGML, PDF, and other published standards is strongly encouraged for the most important content of an ETD.

The second type of standard deals with naming and referring. The NDLTD has adopted two technologies, Persistent Uniform Resource Locators (PURLs)¹⁰ and handles¹¹, to identify the digital objects stored. These technologies will allow each ETD to have a unique and persistent name, a URN instead of a URL, so that regardless of changes in computers and networks, an ETD can always be retrieved when its name is supplied.

The third type of standard deals with metadata, so that each ETD can be adequately described and searched. The NDLTD has two solutions regarding metadata. One deals with MARC, the scheme adopted for library catalog records. Each ETD should have an accompanying MARC record. That record should include all typical fields, as well as the abstract and URNs (in the 856 field). Student authored keywords are strongly encouraged. Depending on local library practice, categorization according to LCSH and/or DDC is encouraged. An experimental effort at Virginia Tech, undertaken in collaboration with OCLC's Scorpion Project staff, is working toward semi-automatic assignment of DDC numbers to assist in this regard.

The second solution regarding metadata is to use the SGML representation. Many of the tags in our tailored SGML Document Type Definition (DTD), which is called ETD-ML (ETD Markup Language), deal with typical metadata requirements of such documents. Conversion tools to map from ETD-ML to MARC, Dublin Core¹², and TEI¹³ will round out the needs for metadata in the most common forms.

Interoperability

For the NDLTD to operate worldwide, it is important that access be easily obtained for works when authors so authorize. There are three key technical aspects of our approach to this problem. First, all NDLTD members are asked to freely share the MARC records for their ETDs with all other NDLTD members. This means that, for example, an NDLTD member could keep a searchable file of all MARC records for all ETDs. While that might be appropriate for universities with heavy demand and that lack high speed network connections, or if some organization is encouraged to run a centralized service, it is much more likely that large MARC record files will exist only at national or regional servers, or at sites providing archival services for a number of institutions.

Second, since MARC records contain URNs in the 856 field, as long as each ETD is accessible when its URN is supplied, searching of the worldwide collection of ETD MARC records will allow access to relevant ETDs. This assumes that access through URNs works, which in turn depends on suitable client software (none is required for PURLs, but plug-ins are needed for handles) as well as servers (initially, PURL and handle servers running at Virginia Tech, but later in a more distributed fashion). Further, ETDs themselves must be accessible from those URNs, which means the URN servers are loaded with URLs for each ETD, and that a university, regional archival service, or other archive makes each ETD accessible when its URL is supplied. Note that if there are access restrictions regarding an ETD, those can be dealt with by a handle server

managing terms and conditions, and/or by each ETD archive enforcing the appropriate restrictions.

Third, regarding searching and browsing, there must be support for handling local, regional, and archival collections, and for federated access. One aspect of that requirement is the matter of search support. At Virginia Tech, for example, the OpenText system is used to allow full-text search of metadata as well as PDF files. While eventually full-text, as well as search on image content and other types of searching, will be supported over the entire collection, initially it is likely that searching of the full NDLTD collection will be based solely on the metadata and/or the MARC records.

This leads to the second aspect of supporting searching namely protocol and system. Here there are a pair of leading approaches. The NCSTRL model can be followed and the Dienst protocol and software that is at the heart of NCSTRL can be utilized. This avenue is being pursued through interoperability tests between the University of Virginia (UVA) and Virginia Tech, based on extensions made at UVA to the Dienst software. In NCSTRL, UVA runs what is called the “central server” and Virginia Tech runs the “backup server”, so building on that experience, it will be straightforward for the NCSTRL model to be fully adapted to the NDLTD.

The other approach to searching and browsing involves Z39.50. A federated search of a number of Z39.50-compliant servers can be undertaken by a suitable gateway system. Software development at Virginia Tech is directed toward that aim, building upon the SiteSearch system of OCLC that has been licensed by OCLC that has been provided to help the NDLTD. All in all, once interoperability tests are completed, access to ETDs from the worldwide distributed collection should be feasible.

In addition to the three technical aspects of interoperability explained above, there are a number of social and organizational considerations (Table 2). Most relate to sharing of data (e.g., summaries from surveys or logs, for joint analysis and comparison), information and experiences. A few bear special discussion.

One key hope for the NDLTD in 1998 is to convene a user group meeting. Representatives from member institutions will be invited to attend a meeting in person or through videoconferencing. Members will share experiences, software, tools, policies and practices.

The other troublesome issue regarding sharing among universities deals with interlibrary loans (ILL). In cases when an ETD is restricted to a campus so that publishers will not deny requests for inclusion in a journal, for example, it is still desirable, nevertheless, for a small number of interested parties to gain access. This might be feasible through a modification of the current ILL system, but technical and administrative details will need to be resolved.

Technical	Social/Organizational
Formats PDF SGML Multimedia Standards	Sharing Data from Logs and Surveys Educational Materials/ Documentation ILL Marc Records Technology
Identification PURL Handle	Communication Listserv User Group Meetings Video Conferences
Metadata Marc ETD-ML	
Searching Z39.50 Dienst OpenText	

Table 2: Interoperability

Scalability

Growth of the NDLTD depends upon its scalability. This matter has four main dimensions.

Authoring

Each student prepares his or her own ETD and submits it to graduate school. This eliminates costly capture and conversion steps.

Quality Control

Each student's graduate committee is responsible for the content and presentation of an ETD, to ensure that appropriate standards of quality are met for that group, department, college, and university. Further, the graduate school, in keeping with prior tradition, ensures a further degree of quality control regarding formatting and other university rules.

University Coordination

Each institution manages its own collection and deals with training, policies, processing, cataloging, and arranging (or implementing) archiving practices. URN servers, search servers, WWW sites for information, and other assistance may be provided locally. In

cases where a university lacks interest, infrastructure, or critical mass in terms of numbers of ETDs, a regional service may handle the workload for a university.

Searching, Browsing, and Access

In reality, the number of ETDs is relatively small. Thus, based on records from UMI, there are only about 1.5 million dissertations on file, and based on records from OCLC, there are only about 3.5 million catalog records for theses or dissertations. Even if there is a growth of 200,000 per year, the collection is still relatively small when compared to what is now managed by large online services, especially if only the metadata or MARC records are considered. However, full-text searching of the full collection in a centralized fashion will require a large amount of resources. If the NCSTRL model, or a Z39.50-based federated search were undertaken, however, parallel searching of about 500 sites (assuming that only large universities run servers, and that national or regional servers are deployed when that makes more sense) would obviate the need for very large central systems. Though this task could be undertaken today, better technical solutions are needed for that type of search to become highly efficient and effective.

PROGRESS

The NDLTD has grown rapidly in its first year. There is widespread interest, support of many commercial groups, as well as some local excitement and local adaptation.

Widespread interest

By summer 1997, the NDLTD had about twenty members (Figure 1). Also, almost a hundred universities had either been visited, heard presentations, or expressed an interest in the NDLTD. It appears likely that membership and interest will continue to grow.

Once Virginia Tech decided to require all students submitting a thesis or dissertation to do so electronically, a wave of publicity spread across the U.S. *National Public Radio*, the *Chronicle of Higher Education*, and the *New York Times* all reported on the initiative. These reports ignored many of the goals of the initiative, and emphasized controversy, but they have led to many additional newspaper articles and familiarity with aspects of the matter by large numbers of listeners and readers.

The NDLTD team has written a series of articles to reach those with scholarly interest in the ETD Initiative. A booklet is being devised to aid students and help them prepare ETDs; this is likely to prove of value to many others interested in electronic document preparation. A symposium series to raise awareness on ETDs is planned for 1998 and should lead to an edited volume on the topic. Further, a sourcebook on ETDs is being prepared as an edited work that may be published by a university press; this could be of even greater value than the above mentioned booklet and reach a broader audience since it will cover a wide variety of issues as well as practical details.

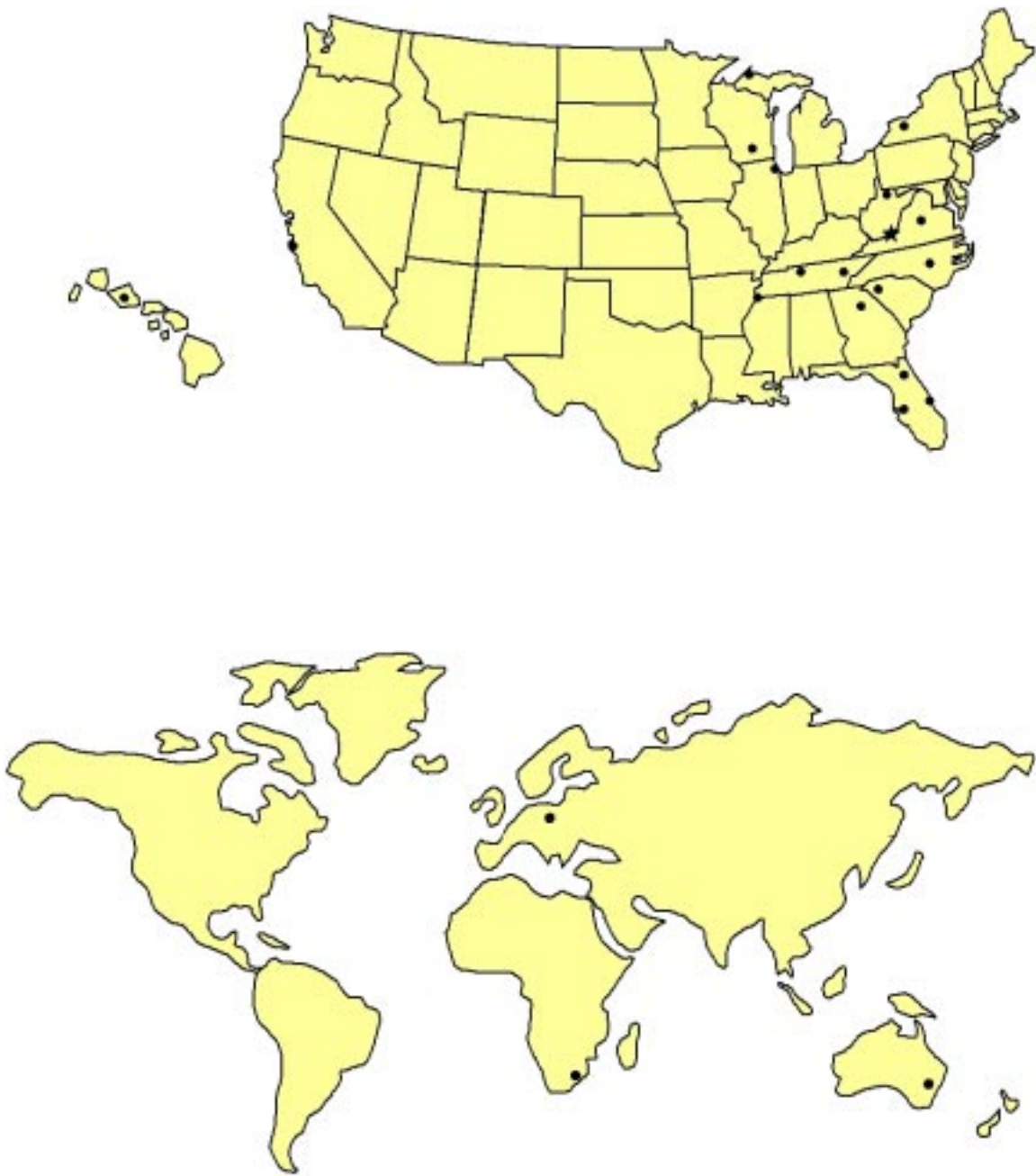


Figure 1: NDLTD Project Membership, October 1997

A variety of other efforts are underway to increase interest and involvement in the NDLTD. Fliers have been prepared suitable for graduate schools and libraries. Videotapes and CD-ROMs afford easy distribution of project materials, as well as sample WWW site contents. Talks at conferences, mailings, email contacts, videoconferences, and other approaches are being pursued to support current members as

well as to encourage others to join. Some of the many materials prepared to aid training and publicity are summarized in Table 3.

Topic	Join the NDLTD Video	NDLTD CD-ROM	ETD Web Pages	Basic Workshop	Advanced Workshop
ETD Advantages	X	X	X	X	X
Joining Procedures	X	X	X		
Word Processor Hints		X	X	X	
LaTeX Hints		X	X		X
ETD-ML		X	X		X
Conversion to PDF		X	X	X	X
Graduate School Submission		X	X	X	X
Multimedia Objects		X	X		X
Copyright Information	X	X	X		
Thesis Formatting		X	X	X	X
PDF Enhancement		X	X		X
Accessing the ETD Library		X	X	X	X

Table 3: Training Materials

Interest in the emerging collection of ETDs also has been encouraging. It appears that the number of accesses to an ETD is well over ten times the number of accesses to paper theses or dissertations. Some works have had an extraordinary number of accesses (Figure 2a). These accesses come from every domain of the Internet (Figure 2b), and from many nations (Figure 2c).

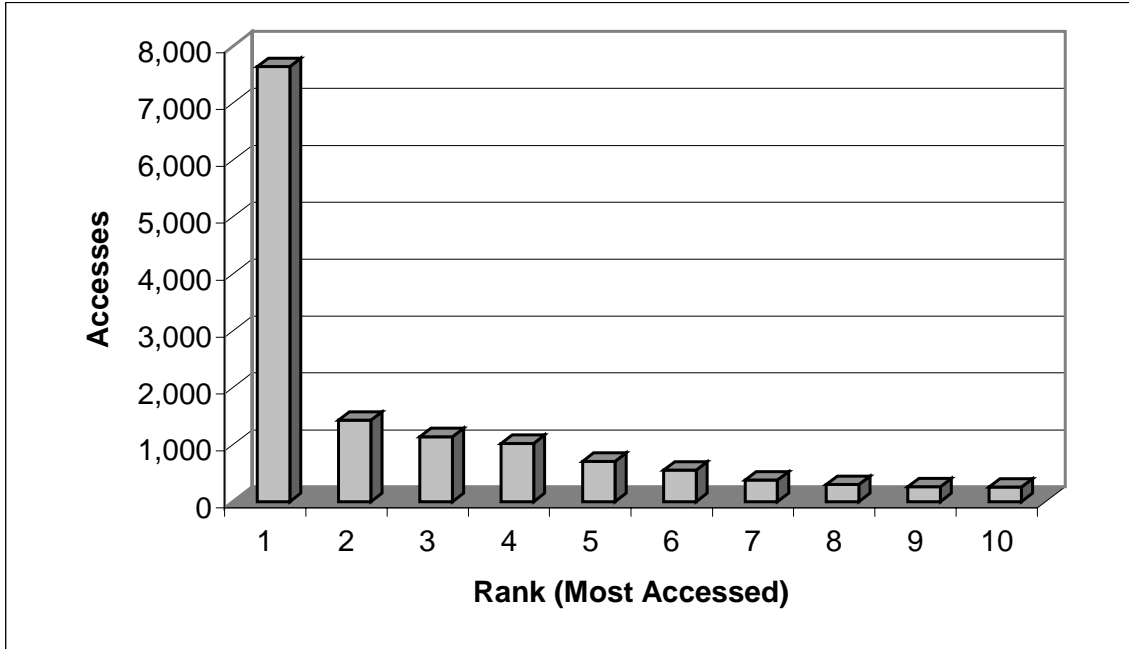


Figure 2a: Accesses by Work

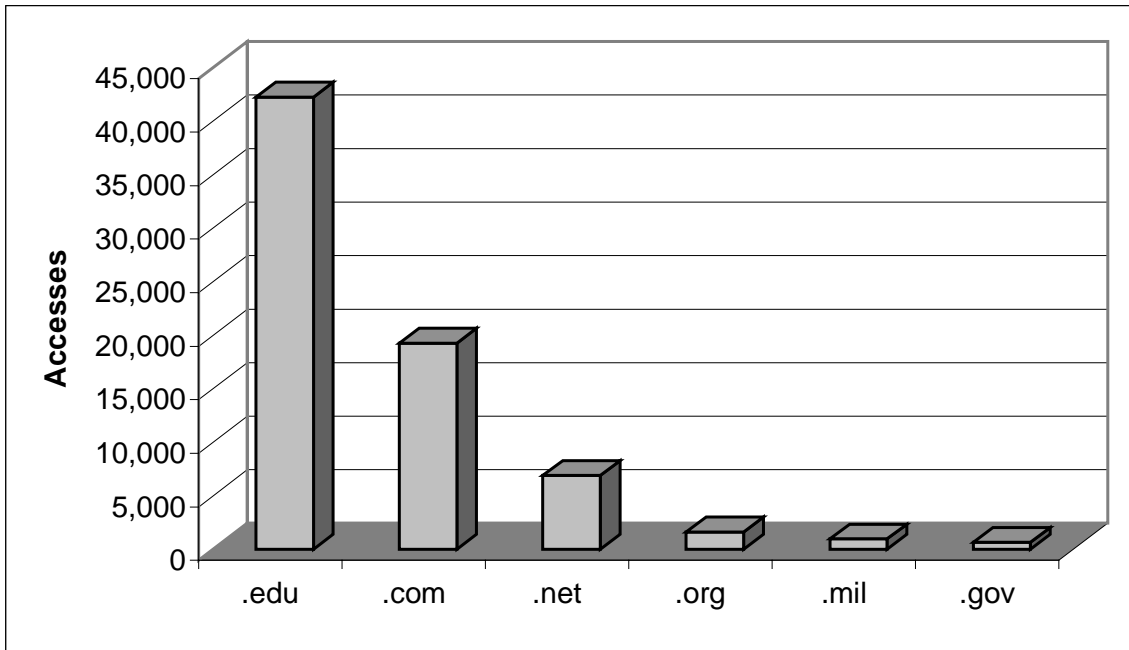


Figure 2b: Accesses by Domain

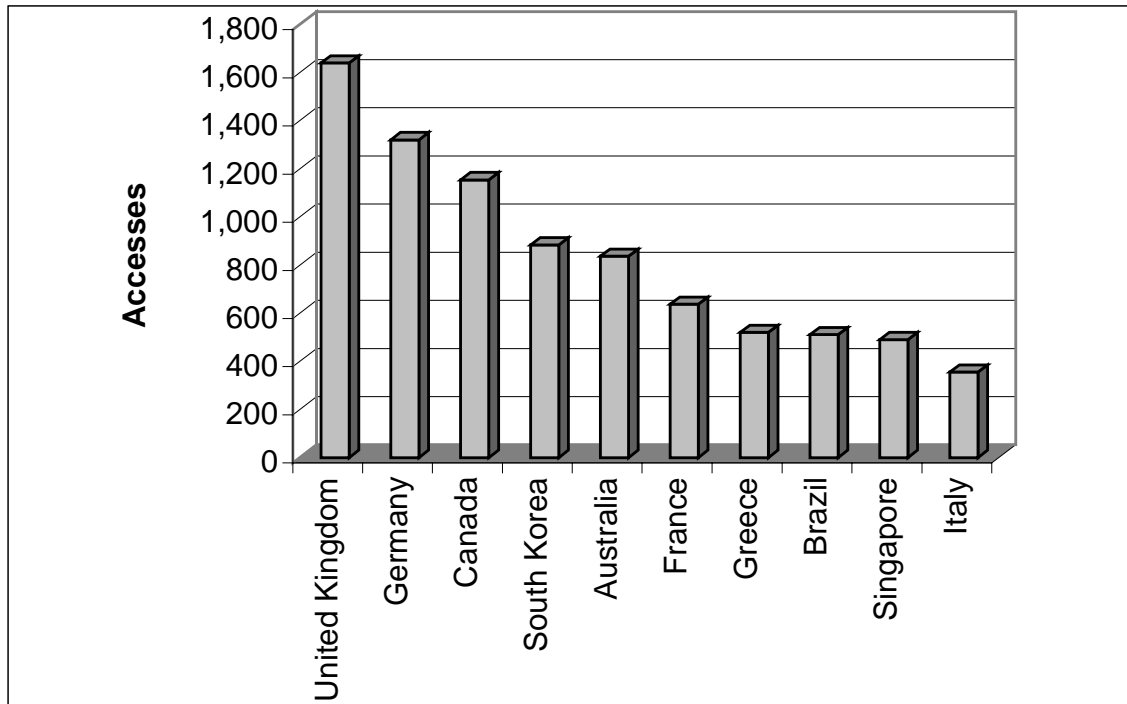


Figure 2c: Accesses by Country (outside U.S.)

Commercial interest

The NDLTD has had considerable corporate support since its inception. Work by UMI to develop a digital library based on its own holdings is strongly encouraged as an aid to the overall project goals. OCLC has provided technical assistance and a free site license for SiteSearch. IBM has contributed \$650,000 worth of equipment to aid the R&D work on digital libraries at Virginia Tech. Adobe and Microsoft have made generous donations of software to aid universities joining the initiative.

Another type of initiative deals with archiving. Both UMI and OCLC early on indicated interest in providing long-term archival services for ETDs. UMI has committed to archiving works it receives. OCLC has considered offering a similar service.

One other commercial effort relating to the ETD Initiative is dissertation.com, which encourages individuals to make their ETDs available through their system, and indicates that royalties will be paid to students depending on access. It is not clear what impact this will have, if free access to ETDs is afforded by universities, or whether students will get more income from royalties than it costs to file their work with dissertation.com. Nevertheless, this commercial effort also encourages students to create ETDs, and increases access to those works. The NDLTD Steering Committee decided at its spring 1997 meeting that a wide variety of university, governmental, and commercial services that support one or more of the goals of the NDLTD should be encouraged so that learning and access are increased.

Local excitement and adaptability

The ETD initiative at Virginia Tech, while generally supported by students, faculty, and administrators, has led to some local excitement. Because of the turmoil in the publishing world (see “Publishing World”), some students and faculty who did not fully understand the issues and the local policies relating to the initiative became concerned¹⁴.

Educational activities increased to deal with the concerns, and the approval form that contained the agreement by which access is specified and authorized was reworked a number of times to reduce faculty and student responsibility and eliminate contention with publishers. This type of campus discussion and refinement of policies has led to tremendous increase in knowledge, on the part of students, faculty, and administrators, of all aspects of electronic document preparation, publishing policies, intellectual property rights, and digital libraries. It is extremely encouraging that universities can quickly develop the infrastructure, policies, practices, and intellectual climate so that the ETD Initiative can flourish.

CONCLUSIONS

The NDLTD has had a very promising first year. It has ambitious goals, hoping to prepare the next generation of scholars for the Information Age. It fits into a complex context, with changes in the publishing world, continual improvement in technologies and support for electronic document preparation, and the emergence of large-scale digital libraries. The strategies adopted by the NDLTD, based on standards, interoperability, and scalability, should lead to continued growth and responsive services. Progress has been rapid, and is reinforced by a number of commercial efforts. We encourage all who read this article to participate in the NDLTD, and to assist in its unfolding.

REFERENCES

1. Fox, Edward A., John L. Eaton, Gail McMillan (1996). “Improving Graduate Education with a National Digital Library of Theses and Dissertations.” Proposal submitted to the United States Department of Education. 1996.
<<http://www.ndltd.org/support/fipse10.pdf>>
2. Fox, Edward, Eaton, J., McMillan, G., Kipp N., Weiss L., Arce E., and Guyer S. “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>>

-
3. Fox, Edward, Hall, R., Kipp N., Eaton, J., McMillan, G., and Mather, P. "NDLTD: Encouraging International Collaboration in the Academy." *DESIDOC Bulletin of Information Technology*, in press, 1997.
<<http://www.ndltd.org/journal/dbit.pdf>>
 4. Licklider, J.C.R. *Libraries of the Future*. Cambridge, MA: MIT Press, 1965.
 5. Schatz, Bruce and Chen, Hsinchuun. Building Large-Scale Digital Libraries. Guest Editors' Introduction, *IEEE Computer*, 29(5):22-26 May 1996.
<<http://computer.org/computer/dli/>>
 6. Chien, Y.T. "Every Person, Every Society, and in Every Part of the Globe." *D-Lib Magazine*, October 1997. <<http://www.dlib.org/dlib/october97/10editorial.html>>
 7. Kahn, Robert E. "An Introduction to the CS-TR Project"
<<http://www.cnri.reston.va.us/describe.html>>
 8. Davis, James R. "Creating a Networked Computer Science Technical Report Library" *D-Lib Magazine*, September 1995. <<http://www.dlib.org/dlib/september95/09davis.html>>
 9. Coombs, James et al. Markup Systems and the Future of Scholarly Text Processing. *Communications of the ACM*, 30 (11): 933-947, Nov. 1987.
 10. Weibel, Stuart, Erik Jul, and Keith Shafer, "PURLs: Persistent Uniform Resource Locators" <<http://purl.oclc.org/OCLC/PURL/SUMMARY>>
 11. CNRI. "Handle System Overview v3.0" <<http://www.handle.net/overview.html>>
 12. OCLC. "Dublin Core Metadata" <http://purl.oclc.org/metadata/dublin_core/>
 13. Plotkin, Wendy and C.M. Sperberg-McQueen. Text Encoding Initiative
<<http://www.uic.edu:80/orgs/tei/>>
 14. Fox, Edward, Eaton, J., McMillan G., Kipp N., Mather P., McGoingle T., Schweiker W., and Devane B. "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>>