**Chapter XXX**

**Networked Digital Library of Theses and Dissertations (NDLTD)**

Edward A. Fox, Gail McMillan, Hussein Suleman, Marcos A. Gonçalves, Ming Luo

## 1    Introduction

This chapter describes the Networked Digital Library of Theses and Dissertations (NDLTD, see www.ndltd.org), as an example of digital library practice. It builds upon discussion in earlier chapters on policy – touching on content, preservation, evaluation, and economics. In the remainder of this section we explain the rationale (Section 1.1), give a library perspective (Section 1.2), and explain our very broad perspective regarding evaluation. In the next section (2) we describe the community served and involved. Section 3 discusses the content, especially at the level of collection, considering its management, size, and access. Section 4 relates this work to the world of scholarly publishing, including perspectives of authors and publishers, considering as well intellectual property rights and preservation. The last section points toward future growth in membership, organization, and services.

### 1.1    Purpose, goals, and objectives

NDLTD was launched in 1996 to enhance graduate education through the deployment of digital library technology. Its goals include to: ensure that graduate students are prepared to function in the Information Age, enhance the expressiveness of theses and dissertations, expand the infrastructure in universities to support institutional repositories, and broaden access to student research worldwide. Its objectives include to: increase graduate student understanding of electronic publishing and digital library concepts and technologies, allow integration of multimedia and hypermedia content and methods when appropriate in theses and dissertations, make it easy for universities to host a digital library of their own (student) works, and allow students worldwide much freer access to much larger amounts of scholarly research undertaken at universities.

### 1.2    Library perspective

Theses and dissertations document the academic heritage of a university and serve as prototypes of scholarly communications from budding researchers and the future professoriate. These documents typically contain a survey of current awareness of published literature, the latest research methodologies, and new findings in research and scholarship.

For decades libraries have stored and circulated this final product of graduate students' educational efforts. With the expansion of the Internet have come dramatic changes in the whole enterprise of research and education including electronic theses and dissertations (ETDs) that now reach outside the realm of the university (i.e., beyond students, faculty, and libraries) into the research community and the world of digital libraries. Theses and dissertations always have been a means of sharing knowledge and culture; doing so

electronically expands the methods and approaches available for learning at distant locations, particularly internationally, and gives students the opportunity to learn about electronic publishing and digital libraries. Educational initiatives like ETDs that target graduate students have the potential to help train future generations of scholars, researchers, and professors. ETDs are one genre within the larger world of electronic publications representing major changes and challenges to established ways of thinking and operating within the academic and research communities.

ETDs provide libraries with the opportunity to improve services and increase accessibility to information for current as well as future users. Libraries can save both time and money by reducing or eliminating many manual processes. Even without benefiting from procedures for deriving the cataloging record from the digital document, because ETDs do not have to be bound, labeled, security stripped, bar-coded, checked out, checked in, shelved, and re-shelved, libraries can save about 73% of the cost of processing paper theses and dissertations (McMillan, 2001b).

A problem avoided by ETDs is the one title / one user limitation of works on paper and microfilm; an ETD is simultaneously accessible to multiple users. ETDs are more frequently accessed than their paper counterparts but increased use does not require additional library staff time. Computer programs "move" submitted works through the approval process to availability, with security/backup copies and archiving at a variety of locations. Since there clearly are so many benefits that derive from ETD programs, we lead into later discussion through the following brief introduction to evaluation.

## 1.3    Evaluation

Digital library evaluation is recognizably an extremely complex and difficult task (Saracevic, 2000; Fuhr et al., 2001). The complexity is mainly due to the inherently inter-disciplinary nature of the field (Fox & Marchionini, 1998; Gonçalves et al., 2003) as well as the competing visions of the different communities involved in the area (e.g., research vs. practice, information science vs. computer science) (Borgman, 1999).

We have adopted a broad and integrated vision of digital libraries in which to consider a number of elements for evaluation of NDLTD, including numerous human, system, and society-centered criteria. Those include membership and collection growth, including considerations about international interest and support, access (information seeking activities, physical distribution); student learning and skills development; worldwide availability of ETDs; and qualitative and economic aspects (e.g., usability and economic impact). Our evaluation instruments include logging accesses, collecting surveys, holding focus groups, and undertaking usability studies of digital libraries. Details of each of these instruments, the metrics they use, and results of the analyses, are presented below, after we consider key aspects of community and content.

## 2    Community

According to our "5S" theory of digital libraries, we should consider five key aspects of such advanced information systems: Societies, Scenarios, Spaces, Structure, and Streams (Gonçalves et al., 2003). The first of these concerns users, teams, collaboration, target audience groups, social concerns, and community issues. In the case of NDLTD, we are

interested in students, faculty (advisors, mentors, examiners), librarians, graduate administrators, researchers, universities, and other institutions supporting ETD activities. The latter two groups directly relate to NDLTD since they make up its members.

## 2.1   Membership

Tables 1-3 show NDLTD membership as of March 2003.  In less than four years, NDLTD has tripled the number of registered members (from 59 members in May 1999). There are currently 176 members: 67 U.S. universities (Table 1), 86 non-U.S. universities (Table 2), and 23 institutions / regional centers / organizations such as UNESCO (Table 3). These numbers demonstrate the growth of global interest in ETDs as international participation, represented as a fraction of the total membership, grew from less than one third in 1999 to one half in 2003.  Also, by early 2002, at least 11 of the registered NDLTD members already had started requiring mandatory submission of electronic theses and dissertations, indicating a very strong commitment to the initiative. (In Tables 1-2, universities currently requiring ETDs are marked with an asterisk.)

**Table 1. NDLTD Membership—US Universities**

| USA Universities (67) | |
|---|---|
| - Air University (Alabama) | - University of Central Florida |
| - Baylor University | - University of Colorado |
| - Boston College | - University of Florida |
| - Brigham Young University | - University of Georgia |
| - California Institute of Technology | - University of Hawaii at Manoa |
| - Clemson University | - University of Illinois, Urbana-Champaign |
| - College of William and Mary | - University of Iowa |
| - Concordia University (Illinois) | - University of Kentucky |
| - Drexel University | - University of Maine* |
| - Duquesne University | - University of Missouri-Columbia |
| - East Carolina University | - University of Nevada, Las Vegas |
| - East Tennessee State University* | - University of New Orleans |
| - Florida Institute of Technology | - University of North Texas* |
| - Florida International University | - University of Oklahoma |
| - Florida State University | - University of Pittsburgh |
| - Georgetown University | - University of Rochester |
| - George Washington University | - University of South Florida |
| - John Hopkins University | - University of Central Florida |
| - Louisiana State University* | - University of Tennessee, Knoxville |
| - Marshal University | - University of Tennessee, Memphis |
| - Massachusetts Institute of Technology | - University of Texas at Austin* |
| - Miami University of Ohio | - University of Virginia |
| - Michigan Tech | - University of West Florida |
| - Mississippi State University | - University of Wisconsin, Madison |

| | |
|---|---|
| - Montana State University | - Vanderbilt University |
| - Naval Postgraduate School | - Virginia Commonwealth University |
| - New Jersey Institute of Technology | - Virginia Tech* |
| - New Mexico Tech | - Wake Forest University |
| - North Carolina State University* | - West Virginia University* |
| - NorthWestern University | - Western Kentucky University |
| - Pennsylvania State University | - Western Michigan University |
| - Regis University | - Worcester Polytechnic Institute |
| - Rochester Institute of Technology | - Yale University |
| - Texas A&M University | |

**Table 2. NDLTD Membership—International Universities**

| International Universities (86) | |
|---|---|
| - Aristotle University of Thessaloniki (Greece) | - Universitat de Barcelona (Spain) |
| - Assumption University of Thailand (Thailand) | - Universitat de Girona (Spain) |
| - Australian National University (Australia) | - Universitat de Lleida (Spain) |
| - Biblioteca de Catalunya (Spain) | - Universitat Oberta de Catalunya (Spain) |
| - Centro University La Salle-UNILASALLE (Brazil) | - Universitat Politecnica de Catalunya (Spain) |
| | - Universitat Politecnica de Valencia (Spain) |
| - Chinese University of Hong Kong (Hong Kong) | - Universitat Pompeu Fabra (Spain) |
| | - Universitat Rovira i Virgili (Spain) |
| - Chung Yuan Christian University (Taiwan) | - Universitat Sbibliothek Munchen (Germany) |
| - Chungnam National U., Dept of CS (S. Korea) | - Universitè Laval (Quèbec, Canada) |
| - City University, London (UK) | - Universitè Lyon 2 (France) |
| - Curtin University of Technology (Australia) | - University Catolica de Brasilia-UCB (Brazil) |
| - Darmstadt University of Technology (Germany) | - University Catolica de Pernambuco-UNICAP (Brazil) |
| - Freie Universitat Berlin (Germany) | - University Catolica de Salvador-UCSAL (Brazil) |
| - Gerhard Mercator Universitat Duisburg (Germany) | - University Catolica de Santos-UNISANTOS (Brazil) |
| - Griffith University (Australia) | - University Catolica Dom Bosco-UCDB (Brazil) |
| - Gyeongsang National University, Chinju (Korea) | - University do Vale do Rio dos Sinos-UNISINOS (Brazil) |
| - Helsinki University of Technology (Finland) | - University of Antioquia (Medellin, Colombia) |
| - Humboldt-Universität zu Berlin (Germany) | - University of Bergen (Norway) |
| - Indian Institute of Technology, Bombay (India) | - University of British Columbia (Canada) |
| | - University of Edinburgh (UK) |
| - Lund University (Sweden) | - University of Glasgow (UK) |
| - McGill University (Canada) | - University of Guelph (Ontario, Canada)" |
| - Nanyang Technological University (Singapore) | - University of Hong Kong* (Hong Kong) |
| - Naresuan University (Thailand) | - University of Hyderabad (India) |
| - National Sun Yat-Sen University (Taiwan) | - University of Melbourne (Australia) |
| - National Taiwan Normal University (Taiwan) | - University of Mysore (India) |

| | |
|---|---|
| - National University of Singapore (Singapore) | - University of New Brunswick (Canada) |
| - Pontificia Universidade Católica de Minas Gerais (Brazil) | - University of New South Wales (Australia) |
| - Pontificia Universidade Católica do Rio de Janeiro (Brazil) | - University of Novi Sad (Yuguslavia) |
| | - University of Pisa (Italy) |
| - Pontificia University Catolica de Campinas-PUC/CAMPINAS (Brazil) | - University of Pretoria (S. Africa)) |
| - Pontificia University Catolica do Parana-PUC/PR (Brazil) | - University of Queensland (Australia) |
| | - University of São Paulo (Brazil) |
| - Pontificia University Catolica do Rio Grande do Sul-PUC/RS (Brazil) | - University of Sydney (Australia) |
| | - University of Tampere (Finland) |
| - Rand Afrikaans University (South Africa) | - University of the Free State (South Africa) |
| -Rhodes University (South Africa)* | - University of Utrecht (Netherlands) |
| - Shanghai Jiao Tong University (China) | - University of Waterloo (Canada) |
| - St. Petersburg State Technical U. (Russia) | - University of Western Ontario (Canada) |
| - State University of Campinas (Brazil) | - Uppsala University (Sweden) |
| - UNESP – Universidade Estadual Paulista (Brazil) | - Wilfrid Laurier University (Canada) |
| | - Xiamen University Library (China) |
| - Universidad de Chile (Chile) | - Yupei University of Science and Technology (Taiwan) |
| - Universidad de las Amèricas Puebla (Mèxico) | |
| - Universidad Politecnica De Cartagena (Spain) | |
| - Universidade Federal Fluminense (Brazil) | |
| - Universitat Autonoma de Barcelona (Spain)* | |
| - Universitat d'Alacant (Spain) | |

**Table 3. NDLTD Membership—Institutions**

| Institutions (23) | |
|---|---|
| - British Library | - MathDISS International |
| - Cinemedia | - National Documentation Centre (NDC, Greece) |
| - Coalition for Networked Information | - National Library Of Canada |
| - Committee on Institutional Cooperation | - National Library of Portugal |
| - Comunidate Virtual de Aprendizagem Da Rede de institucoes Catolicas de Ensino Superior | - Office of Scientific and Technical Information (Department of Energy) |
| - Consorci de Biblioteques Univ. Catalunya | - OhioLINK |
| - Diplomica.com | - OCLC |
| - Dissertationene Online | - Organization of American States (OAS) |
| - Dissertation.com | - SOLINET |
| - ETDweb | - Sudanese National Electronic Library (Sudan) |
| - Ibero-American Sci. & Tech. Ed. Cons. (ISTEC) | - UNESCO |
| - Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) | |

## 2.2 National Projects

NDLTD is a worldwide initiative. It has worked to expand awareness of the benefits of ETD initiatives at universities around the globe. In the following we give a brief description of five international projects, which have been influenced by or collaborated with NDLTD. Others are evolving, e.g., in Spain and China, but due to lack of room we cannot describe them here.

**USA.** Work on ETDs began with a meeting in Ann Arbor, Michigan, USA on November 1987 when UMI invited a number of universities and companies to consider how SGML (Standard Generalized Markup Language, an ISO international standard and parent of both HTML and XML) might support a move towards electronic versions of dissertations. Over the next several years, Virginia Tech, working with SoftQuad, developed an SGML Document Type Definition and a small number of SGML representations of dissertations. However, it did not seem feasible to spread this methodology to other institutions until Adobe developed the Portable Document Format, PDF, and the Acrobat family of supporting tools. Once that occurred, Virginia Tech began working with UMI, the Council of Graduate Schools, and the Coalition for Networked Information to support ETD efforts. A series of meetings, each involving roughly ten universities, allowed key concepts to develop and interest to expand. In 1995, the Southeastern Universities Research Association awarded a grant to Virginia Tech to develop and disseminate the concept in the region. In 1996, the US Department of Education awarded a grant to Virginia Tech to expand the effort nationwide. Thus the NDLTD was born, and a base of interest grew in the USA. While this has typically been done at the level of individual universities, in Ohio a program developed, supported by OhioLink, to allow engagement by all public and private colleges and universities. Growth of NDLTD in USA was further encouraged by a series of annual conferences. Though international in scope and attendance, it is only with the sixth (May 2003) in the series, that the event is outside the USA, in Berlin, Germany.

**Germany.** After the US, the first serious financial support for ETD activities was by the German government, reflecting the keen interest in dissertations in that country. 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 early 1998. It was completed in March 1999, with a conference held in Jena, Germany. DGF funded a second research and development project, March 1999 to March 2000, at a level of EU 300,000, placing heavy emphasis on collaboration with libraries and university computing centers. Among the learned societies involved in these projects were chemistry, computer science, education, mathematics, and physics. Participants in the second proposal also included five German universities, computing centers, libraries, and the German National Library (DDB). Please see http://www.educat.hu-berlin.de/diss_online/englisch/index1e.html.

**Australia.** Seven institutions in Australia, led by the library at the University of New South Wales, began collaborating in 1998/99 to accept electronic theses from postgraduate students. This "Australian Digital Theses Project" (http://adt.caul.edu.au/) has standardized on SGML and PDF as document formats. The collection's oldest work dates back to 1968.

The ADTP has led to a national program under the auspices of the Australian National Library. It is expected that Australia will be the first nation with comprehensive support for ETD efforts at all universities.

**India.** The Vidyanidhi project (Urs & Raghavan, 2001), based at the University of Mysore and sponsored by India's National Information System for Science and Technology (NISSAT), is a national effort to create, maintain, and provide network access to a digital library of Indian theses. It is a direct consequence of the initiatives identified in India's Information Technology Action Plan. The impetus has come primarily from a policy initiative that makes it mandatory for all universities in the country to host "every dissertation/thesis on a designated Web site." Vidyanidhi is intended to demonstrate the utility of digital library technologies in maintaining as well as enhancing both access to, and visibility of, Indian academic research. Vidyanidhi will eventually emerge as a distributed input and database environment for the ETDs digital library. It has begun, however, as a centralized repository facilitating network access. This is necessary and important for the purpose of evolving a national consensus and agreement on all relevant issues and standards. However, the goal—to assist the end-user community in obtaining access to one of the most neglected and under-utilized of all resources—remains irrespective of the structure and technology that ultimately emerges.

**Brazil.** The Brazilian Digital Library in Science and Technology Project (BDL), developed by the Brazilian Institute for Scientific and Technical Information (IBICT), stresses the impact of the Web on publishing and communication in science and technology and also on information systems and libraries. The two main objectives of the BDL project are: 1) promoting electronic publishing of different full text materials (e.g., theses, journal articles, papers in events, "grey" literature) by the Brazilian scientific community to amplify their national and international visibility; and 2) achieving interoperability among those heterogeneous electronic information resources available on the Web. The project also provides consulting services to help institutions implement their own ETD programs as well as the Open Archives Initiative protocol to become data providers (see Section 3.1). In addition to supporting the Dublin Core, the BDB uses a specific metadata format to meet the requirements of Brazilians institutions and funding agencies. The ETD portion of the BDL intends to collect ETD metadata from all Brazilian institutions of higher education into a Union Catalog and to run services on top of the catalog. Interest in BDB is growing, as is involvement in NDLTD; in March 2003 all 10 of the Catholic universities in Brazil joined. This should help expand content, which is the topic of the next section.

## 3   Content

At the heart of NDLTD is an electronic thesis or dissertation (ETD). These works typically are described using a metadata standard, such as MARC21 or ETD-MS (Atkins et al., 2001). The latter is based on the Dublin Core standard, and was developed by the NDLTD Standards Committee after several years of international discussion. Thus, NDLTD has encouraged improved management of ETD content worldwide.

## 3.1   Management

The ETD collections managed by NDLTD members have traditionally been independent of one another. Thus, in most cases, an ETD can be located by an end-user only if the archive containing the ETD is consulted directly. This is far from the ideal case of an end-user searching through all available ETDs through a unified interface.

To get closer to this objective, NDLTD has adopted the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) (Lagoze and Van de Sompel, 2002) as a means of accumulating metadata from all member sites into a merged collection. The OAI-PMH is an application-layer network protocol to transfer XML-encoded metadata records from one machine to another. Core features of the protocol exploited by NDLTD include the ability to obtain only recently updated records on a periodic basis and the support for multiple metadata formats. The latter has enabled the use of MARC21 and/or ETD-MS for theses and dissertations, in addition to the required Dublin Core format.

All NDLTD members are encouraged to support the OAI protocol, thereby enabling access to their ETD metadata using a simple and open standard. NDLTD maintains a Union Catalog (i.e., an OAI archive) that harvests this metadata from each collection on a periodic basis and republishes it as a single merged collection. This merged collection is then used by service providers such as VTLS (VTLS, 2002) and the ODL-based ETD Union Catalog (Suleman, 2002) in order to supply global discovery services, such as a cross-collection search engine, to users.

## 3.2   Size

The number of ETDs across the NDLTD universities/institutions has grown at a rapid pace: from a few dozen at Virginia Tech in 1996, to 4,328 ETDs at 21 institutions in March 2000, to 7,268 ETDs at 25 institutions in July 2001, and 13,724 at 35 member institutions in November 2002. Table 4 shows a breakdown of the numbers of ETDs as of November 2002, organized by member institution. The statistics were collected from three sources: 1) a count of the number of ETD metadata records in the NDLTD official Union Catalog as of November 2002, harvested from member institutions that already implement an OAI interface (represented by the code UA in the table column "source of information"); 2) an on-line survey conducted by Gail McMillan in July 2001, which represents only those institutions that responded to the survey by that time (code SV); and 3) an independent experiment, performed in May 2002, on crawling and extracting metadata from ETD web sites (Calado et al., 2003) (code CW). For overlapping institutions (i.e., those that appear in more than one source of information), we chose the maximum number (which in most of the cases accounts for the most recent numbers).

**Table 4. NDLTD collection size**

| University/Institution | ETD Collection size | Source of Information (maximum) |
|---|---|---|
| ADT: Australian Digital Thesis Program (Australia) | 238 | SV |

| | | |
|---|---|---|
| University of Bergen (Norway) | 45 | SV |
| California Institute of Technology | 364 | UA |
| Concordia University | 3 | CW |
| Consorci de Biblioteques Univ. de Catalunya (Spain) | 151 | SV |
| East Tennessee State University | 106 | SV |
| Humboldt-University (Germany) | 439 | CW |
| Louisiana State University | 352 | UA |
| Mississippi State University | 33 | SV |
| MIT | 62 | SV |
| National Sun Yat-Sen University (Taiwan) | 1786 | CW |
| North Carolina State University | 301 | SV |
| Ohio Link | 932 | CW |
| Pennsylvania State University | 83 | SV |
| Pontifical Catholic University (PUC-Rio) (Brazil) | 90 | SV |
| Gerhard Mercator Universitat Duisburg (Germany) | 412 | UA |
| Rhodes University (South Africa) | 134 | CW |
| Technische Universität Dresden | 18 | UA |
| Universitat Politecnica de Valencia (Spain) | 189 | SV |
| University of British Columbia | 2 | UA |
| University of Florida | 174 | SV |
| University of Georgia | 121 | SV |
| University of Iowa | 6 | SV |
| University of Kentucky | 30 | CW |
| University of Maine | 27 | SV |
| University of North Texas | 337 | SV |
| University of South Florida | 40 | UA |
| University of Tennessee | 12 | SV |
| University of Tennessee, Knoxville | 28 | SV |
| University of Virginia | 619 | CW |
| University of Waterloo | 105 | CW |
| Uppsala University (Sweden) | 1711 | UA |
| Virginia Tech | 3646 | UA |
| West Virginia University | 1006 | SV |
| Worcester Polytechnic Institute | 122 | CW |
| TOTAL | 13724 | |

These statistics do not take into account scanned theses and dissertations, which make up a substantial portion of the total NDLTD collection. There are roughly 26 ETDs from scanned works at the New Jersey Institute of Technology, 150 at the University of South Florida, 5581 at MIT, and 12000 at the National Documentation Center in Greece. This

sample gives a total of 17,751 scanned theses and dissertations; which are in addition to the 13,724 "born digital" ETDs. Thus, we know of over 30,000 ETDs and are sure that there are thousands of unreported ones at other institutions, including institutions not in the Union Catalog or covered by the crawling experiment. Furthermore, we note (as per the discussion in Section 4.1) that the numbers given represent only the "world-accessible" works at universities, and ignore the thousands of others in controlled collections (e.g., hundreds of thousands at companies like UMI/ProQuest, or at universities in cases where patent issues are temporarily of concern).

## 3.3 Access

Figures 1-4 highlight the quantitative and qualitative aspects of access to ETDs. Figures 1 and 4 deal with the Union Catalog described above, and so only reflect about a dozen sites. In Figure 1 we compare access by the two main service providers, VTLS running its Virtua software, and Virginia Tech running its Open Digital Library prototype software (ODL). Table 5 gives more details. Figure 4 considers only the ODL service. In contrast, Figures 2 and 3 deal only with the Virginia Tech collection, since we have the access logs available.

**Table 5. Access Statistics for Two Services Built Atop NDLTD Union Catalog**

|  | NDLTD - VTLS Virtua (6/02 – 12/6/02) | ODL-based Prototype Service (10/00 – 12/6/02) |
|---|---|---|
| Successful requests | 241,473 | 883,151 |
| Average successful requests per day | 1,323 | 1,121 |
| Distinct hosts served | 12,820 | 15,782 |
| Data transferred (Gb) | 4.21 | 10.74 |
| Average data transferred per day (Mb) | 23.61 | 13.96 |

Figure 1 demonstrates that use of the Virtua system keeps increasing. As with all systems serving primarily a university community, access counts vary widely from month to month, depending on academic calendars. We also note that the ODL service, which came along much later than the Virtua service, is quite popular, with access counts also increasing rapidly.

Figure 2 provides details of access to the Virginia Tech collection, over the prior 3 years, indicating relative counts for countries. It is clear that the access counts have consistently increased from year to year. It appears that countries with significant Internet infrastructure have among the highest counts, but that there is a considerable spread in interest worldwide. Figure 3 presents the same type of result, focusing on Internet domains

as opposed to countries. As expected, the "edu" domain is most popular, but there also is interest from "com" and "net".

Finally, Figure 4 gives a more functional view of accesses, comparing counts for the various functions. Most accesses involve a request for a metadata record for some ETD. Next in popularity is search, followed by the system harvesting to support the Union Catalog. Other functions include showing all available metadata forms, indexing, and browsing.
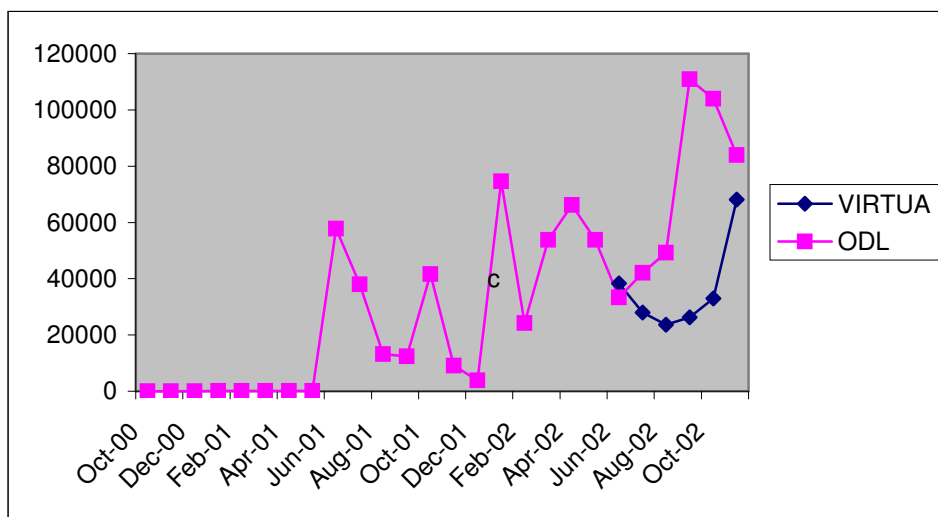


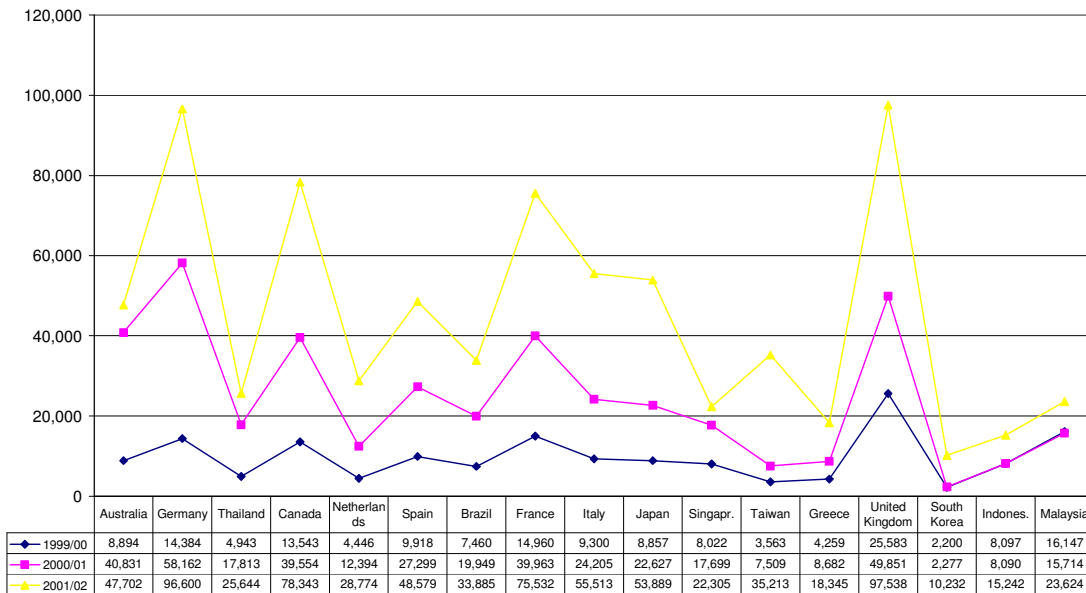**Figure 1. Monthly access to NDLTD Union Catalog (searching + browsing, 10/00-11/02)**



| | Australia | Germany | Thailand | Canada | Netherlands | Spain | Brazil | France | Italy | Japan | Singapr. | Taiwan | Greece | United Kingdom | South Korea | Indones. | Malaysia |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1999/00 | 8,894 | 14,384 | 4,943 | 13,543 | 4,446 | 9,918 | 7,460 | 14,960 | 9,300 | 8,857 | 8,022 | 3,563 | 4,259 | 25,583 | 2,200 | 8,097 | 16,147 |
| 2000/01 | 40,831 | 58,162 | 17,813 | 39,554 | 12,394 | 27,299 | 19,949 | 39,963 | 24,205 | 22,627 | 17,699 | 7,509 | 8,682 | 49,851 | 2,277 | 8,090 | 15,714 |
| 2001/02 | 47,702 | 96,600 | 25,644 | 78,343 | 28,774 | 48,579 | 33,885 | 75,532 | 55,513 | 53,889 | 22,305 | 35,213 | 18,345 | 97,538 | 10,232 | 15,242 | 23,624 |

**Figure 2. Access to VT ETD collection, by country, over 3 years.**
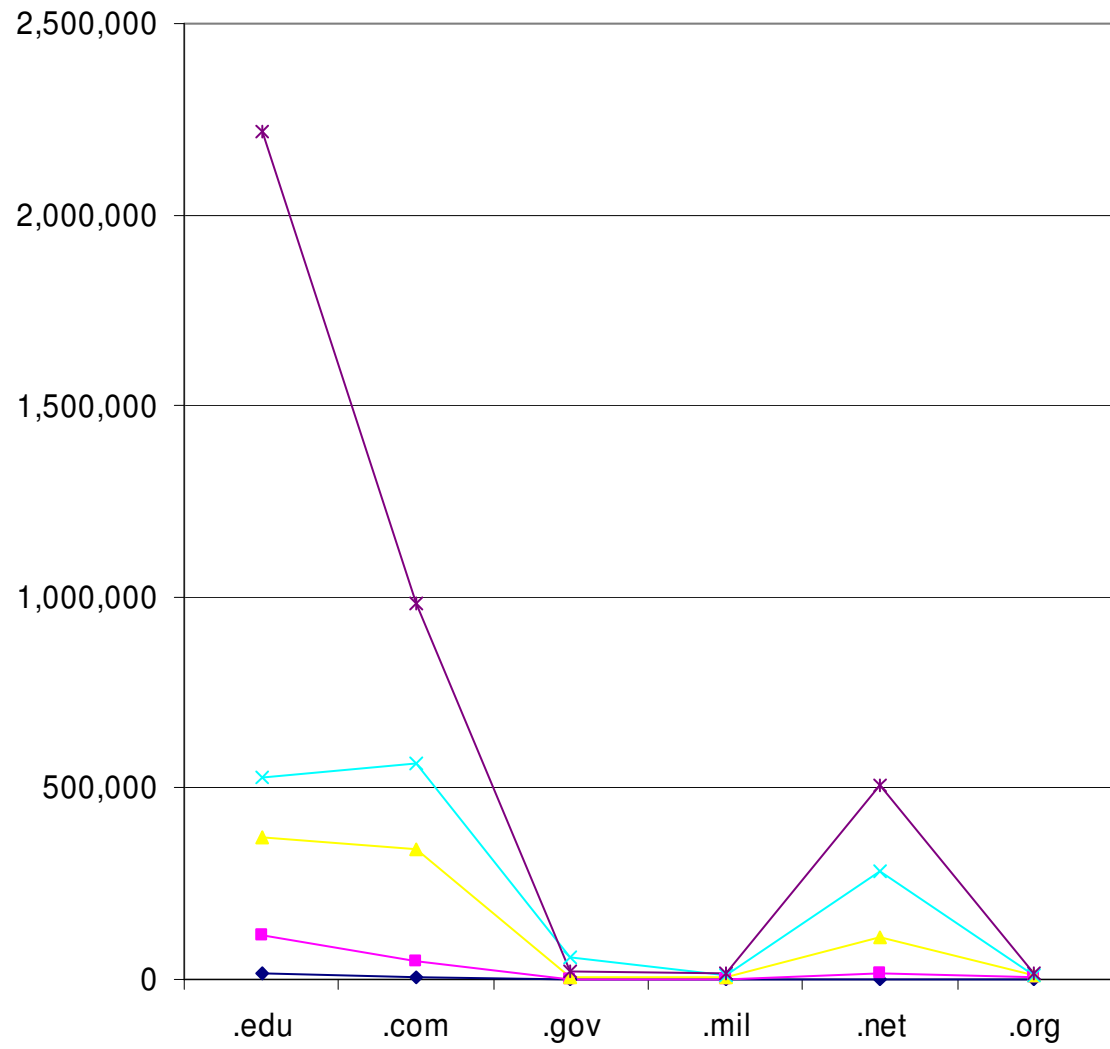
NDLTD for UK DL

**Figure 3. Access to VT ETD collection, by domain, over 4 years.**
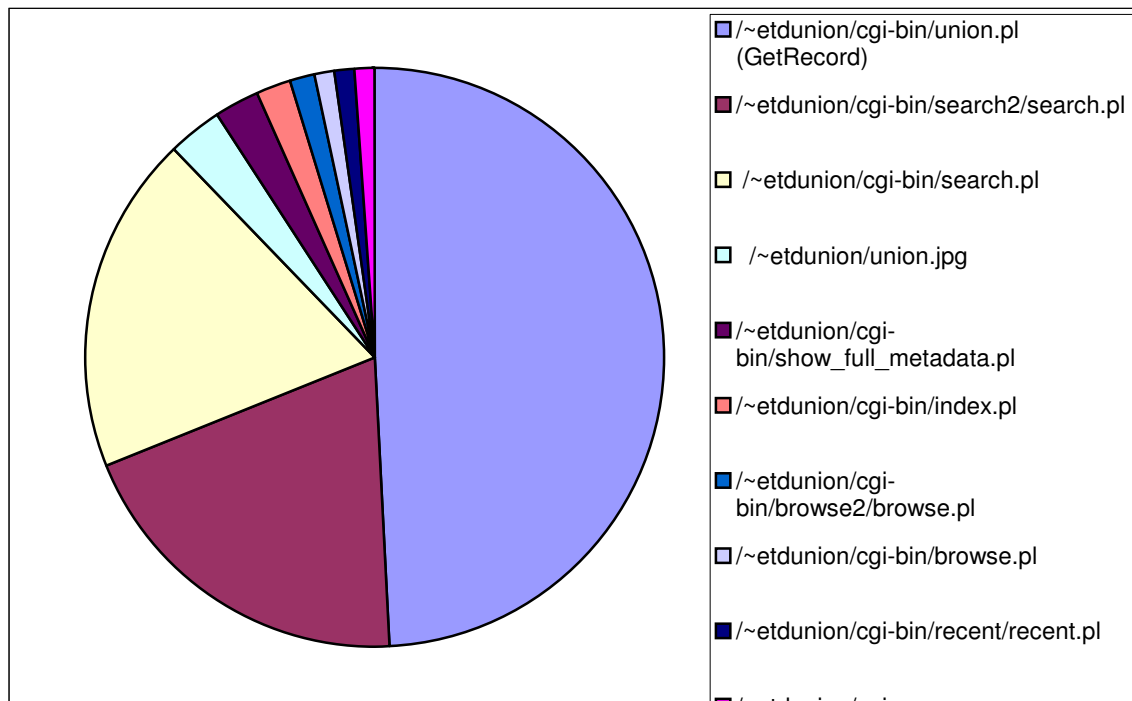
**Figure 4. Most accessed operations, by number of requests (ODL - Union Archive site)**

## 4    Scholarly Publishing Perspectives

In this section we look at ETD issues more from a policy perspective, considering the views of key stakeholders. We also consider author controls on access, handling of intellectual property rights, and preservation in the long term.

Most universities allow their students to make choices concerning who will have access to their ETDs. With a paper based approach, interest and accesses were low, so libraries did not generally have to restrict access to their collections. When the Graduate School withheld access it was primarily to protect patent applications, which were small in number. Today, with ETDs, in part because of the intense interest in ETDs, authors may select from various levels of accessibility (see next subsection), ranging from completely available to entirely restricted/hidden. NDLTD encourages that restrictions should exist for specific and limited periods of time, so that works become accessible as soon as appropriate.

### 4.1    Levels of access

*Unrestricted access* means that readers do not encounter any constraints on Internet access. Libraries providing ETDs with unlimited access include bibliographic data in their online catalogs with links to the digital documents. Many also provide a variety of other access points. The majority of authors at NDLTD member institutions select unrestricted

access for their ETDs, tacitly supporting a timely, easy, and inexpensive means of sharing information.

The option to *partially restrict access* creates the comfortable environment that many feel most closely parallels having bound theses and dissertations in the library. Similar to the situation for various online article databases that libraries license for their university communities, in this case ETD readers can be limited to the author's university and to on-campus library users. Many consider this to be the easiest way to transition from library-only to worldwide access. This level of access is often selected to comply with publishers who threaten to reject journal submissions related to an ETD because of considering any online work published.

An advantage of restricting access to campus is that everyone associated with the originating university can have access from a variety of locations both inside and outside the library at all hours, and when someone else is also looking at the same ETD. Members of the author's university are recognized by their computers' IP addresses or their university-assigned user IDs, like a closed-stacks library where users must show university identification to gain access. InterLibrary Loan will receive requests for restricted access ETDs; this option should only be allowed if authors allow these files to be shared through typical library lending and delivery services beyond the author's university.

Prolonging restricted access to their university-based research and scholarship, should be the responsibility of the authors. Monitoring restricted access can be accomplished programmatically. The challenge arises in determining when to move the document from restricted to worldwide access, e.g., after some maximum time period, of say 1-5 years.

*Fully restricted access* denies all users access to the ETD. Usually there is no information about the ETD in the library catalog or the ETD database. Generally, only the authors and their committees and their family members will know that this document has been completed, along with the Graduate School or approving unit of the university.

Whether completed in paper or electronic formats, there are conditions that require theses and dissertations to be entirely withheld from public access, particularly during US patent applications. This protection is easily provided in the online environment, but when this option is too readily available, many authors will choose it out of fear of the unknown rather than as a result of need. We know from survey data collected from Virginia Tech's authors that many faculty advisors recommend restricting or withholding access to reduce problems for the future generation of faculty of getting published. They fear that traditional scholarly or commercial publishers will not accept a manuscript derived from the whole or part (e.g., chapter, data, appendix, etc.) of an Internet-accessible work. Perhaps they should reconsider this matter; constraining ETD access to their university communities may serve to transition from library-only to worldwide access.

## 4.2   Authors surveys

Comparing the survey responses from the first half of fiscal year 2002/03 and all of fiscal year 1999/2000 reveals some changing attitudes and practices among graduate students submitting ETDs at Virginia Tech.

Answers to questions that arise while preparing one's ETD are generally provided by the same people/groups, over time. We see little difference in the numbers for 2002/03 as opposed to 1999/00. The largest number still consult their friends (17.7% and 20.2%) while

relatively few (though increasing slightly from 1999/00) consult their committee members (6.1% and 4.8%), or the New Media Center (NMC) staff (6.5% and 4.6%). Students found the NMC staff more helpful in 2002/03 (24%) than in 1999/00 (19.8%) and fewer found them not helpful, 1.8% vs. 3%. The VT ETD web continues to be a useful source of information, used by 88% in 1999/00 and 90.6% in 02/03.

Far fewer students attend ETD workshops to learn about the initiative and to secure basic training regarding ETD preparation – 5.8% vs. 11.6% – but those who do, continue to find them useful. The 2002/03 survey respondents found the workshops 3.5 times more useful, while in 1999/00 they found the workshops to be more than 5 times more useful.

According to survey respondents, they found it less difficult than they expected to create a PDF file, down 7.6% to 17.5% of the 2002/03 respondents. Students' expectations that submitting at ETD would be difficult dropped over 5% between 2002/03 and 1999/00, while there was a 6.5% increase in the number that found online submission "much less difficult."

ETDs can be submitted from anywhere. There was more than a 13% increase in the number of students submitting from their off-campus workplaces between 2002/03 and 1999/00. Slight increases in remote submissions were also noted from Virginia Tech's Northern Virginia Center (2.1%) as well as from other universities (3.4%).

Graduate students consult Virginia Tech ETDs, though slightly fewer in 02/03 (74.4%) than in 99/00 (77.9%). In 2002/03, 26.8% consulted 3-5 ETDs, while in 1999/00 only 18.7% consulted this many.

After completing their studies and successfully defending their ETDs, authorship is an expectation of most if not all graduates who earned doctorates. (The survey respondents are not separated, and included both masters and doctoral candidates.) More than 80% plan to publish among the 2002/03 group; conference proceedings will be the venue for 20.74% while 2.36% plan to publish books. About the same percentage each year plan to publish journal articles, ~50%. There was a dramatic increase in the percentage that are uncertain about publishing anything, 15.35% in 2002/03 compared to 0.1% of the 1999/00 respondents.

When asked why they restricted access, the surveys reveal that slightly fewer faculty as well as many fewer publishers are giving this advice. In 2002/03 students said 44.6% of the faculty advised them to restrict access to their ETDs, while in 1999/00 47.6% said they did. However, there was a 5.3% drop in the number of publishers giving this advice, down to 2.9% from 8.2% of the survey respondents.

## 4.3   Intellectual property rights

At the same time that authors select the level of access, they should give the library and the university explicit permission to store and provide access to their works. Working with ETDs provides an opportunity for all participants in this initiative to become familiar with copyright laws. For example, libraries can take the opportunity to document their standard practices and to adapt them for their digital libraries, asking authors to formally permit their libraries to preserve, store, and provide access—just what libraries have done without permission with traditional theses and dissertations. As an aid to universities, we offer the following example statement that is in use at Virginia Tech

### 4.3.1 Sample ETD copyright statement

> I [the author] hereby grant to [the institution] and its agents the non-exclusive license to archive and make accessible under [specified conditions] my thesis, dissertation, or project report in whole or in part in all forms of media, now or hereafter known. I [the author] retain all other ownership rights to the copyright of the thesis, dissertation, or project report. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

> In addition, I hereby certify that, if appropriate, I have obtained written permission from the owner(s) of third party copyrighted matter to be included in my thesis, dissertation, or project report, allowing distribution [as specified].

### 4.3.2 Rights of authors

Whether authors are creating electronic or paper theses or dissertations, it does not change their moral or legal rights and obligations. While university policies vary, it is frequently the custom that the person who creates a work is the owner of the copyright. Therefore, the (student) authors of electronic theses or dissertations are the copyright holders and own the intellectual property, their ETD. It surprises some faculty that, though advisors may have provided the research dollars (e.g., through grants) that funded the research and though they may have provided the laboratory and equipment that graduate students used to conduct the research, the author of the ETD is the copyright holder. Within the United States, this means that authors decide how their works will be reproduced, modified, distributed, performed and displayed in public. Out of courtesy, however, students are encouraged to discuss the matter of access level with their advisor and mentors, so as to promote goodwill and to expand understanding of copyright matters.

## 4.4 Publishers

Publishers, who in some cases early in the life of the ETD initiative threatened the future publications of ETD authors (e.g., American Chemical Society discouraged making ETDs world-accessible), have generally modified their policies. Three independent surveys of publishers showed their willingness to consider articles from ETDs on par with articles from paper theses and dissertations. A large and growing number of publishers readily acknowledge that articles from ETDs are derivative works that can be published in academic scholarly journals. Publishers like Elsevier Science have openly supported the ETD initiative from its earliest days.

The three surveys of scholarly journal publishers were conducted in 1999, 2000, and 2002. They revealed that the overwhelming majority of publishers are not deterred from considering articles from ETDs for their publications. These surveys collected data using Virginia Tech's Digital Library and Archives online survey database to gather and report responses. See the survey data at http://lumiere.lib.vt.edu/surveys/results/.

Joan Dalton (librarian, University of Windsor) surveyed publishers and editors of scientific journals about their policies and attitudes towards ETDs; Dalton reported her findings at the annual ETD conference in 2000 held at the University of South Florida's Tampa campus. Dr. Nancy H. Seamans (librarian as well as graduate student in Instructional Technology at Virginia Tech) surveyed social science and science-and-technology studies journal editors and publishers in late 2000. Bobby Holt (a graduate student in History at Virginia Tech) surveyed humanities book and journal publishers in early 2002. The majority of Seamans's and Holt's respondents were not-for-profit publishers.

In 1999 Dalton found that 86% would publish works derived from ETDs (Dalton and Seamans, 2003). Seamans reported that 82% of her survey respondents in early 2001 answered that they were willing to accept articles from ETDs (Dalton and Seamans 2003). Both Dalton and Seamans reported that there was more a perception, than actual evidence, of a problem (McMillan, 2001a).

Respondents to Holt's early 2002 survey gave conflicting information, though 87% responded that their policies do not specifically refer to works electronically accessible on the Web and 44% responded that editorial policies had not yet been set on this issue. 23% said that editorial policy included ETDs "widely available through a Web-based archive" as prior publication while they commented "…this publication would not rule out consideration of the manuscript" and "a dissertation should be revised before it is submitted to a publisher for consideration." 50% of his respondents also answered, "manuscripts derived from Web-based dissertations are considered on an individual basis." Only 3% said "Under no circumstances. Manuscripts derived from research made widely available via the Web are considered previously published."

Among all the surveys, this response describes the overwhelming opinion:

> …the issue is not whether the material was previously accessible in the print or electronic form of a dissertation. What we look for are works that if they began as dissertations have been significantly revised in such a way that they now represent legitimate book manuscripts derived from mature research and writing.

John Eaton, Dean at Virginia Tech's Graduate School, surveyed graduate student alumni in 1998 and 1999 and asked about publishing articles derived from their ETDs. ALL of those who had successfully published reported that they did not have any problems getting articles accepted due to their theses or dissertations being online and readily available on the Internet.

Keith Jones from Elsevier stated publicly and emphatically at the ETD conference in 2001 that his company encourages its authors to link their articles in Elsevier journals to their personal Web sites and also authorizes their academic departments to provide such links. Jones reported that Elsevier understands the importance of getting new authors such as graduate students to publish in his company's journals early in their careers because then they are likely to continue to publish there. The fact is, for many, publishing in an Elsevier journal is an important source of academic validation, and the subsequent availability of those articles from other non-profit and educational sources is not a threat. Several

publishers have similarly attested that ETDs are not published works, in a variety of sites, such as http://www.ndltd.org/publshrs/index.html.

## 4.5    Preservation

Academic departments determine the quality of the work of their students, while the individual thesis/dissertation committees approve the student's work on its own merits. Often the Graduate School oversees mechanical considerations, the purpose of which is to provide a degree of uniformity, to assure that each thesis or dissertation is in a form suitable for reading and/or viewing online, and that it can be preserved. The University Archives at Virginia Tech ensures long-term preservation and access to these records of graduate students' research.

The best way to preserve electronic information is to keep it alive and continuously used in multiple locations. As soon as it is taken offline and not used (i.e., forgotten), it will be difficult to retrieve the media that produced it and made it accessible. With digital materials libraries give access and simultaneously prolong the life of the work, ensuring the durability of the present through stable media. Prior to ETDs joining the library's collection of traditional theses and dissertations, it is appropriate to announce a commitment to maintain these online information resources for long-term access. A library's Special Collections Department and/or its University Archives are often responsible for storing and preserving theses and dissertations. Typically, they will create and document parallel standards, policies, and procedures for electronic theses and dissertations.

### 4.5.1    Formats

The library must strive to ensure that particularly its unique electronic resources such as ETDs will be available indefinitely. To keep ETDs reader-friendly and to retain full access will mean migrating current formats to new standard formats not yet known. This is why standard formats should be the only files accepted.  Migration may be done through the collaborative efforts of the various computing resources of a university though the library maintains the submission software, the database of ETDs, and the secure archive.

The life of an online dissertation in PDF format is said to be at least 100 years according to OCLC research scientists. In addition, the Acrobat Reader is an open source program; systems programmers could potentially create their own PDF readers if the need arose. Still, many feel very uncomfortable about preserving and archiving ETDs because they do not have the historical evidence that paper and – to a certain extent –  microfilm have. Nothing can change this but the passage of time.

Other multimedia found in ETDs such as video (e.g., .qt, .mov) and sound files (e.g., .aif, .wav) have shorter life expectancies. It is better to use international standards, like MPEG and JPEG. Further, when these and other file formats (as well as HTML links) accompany PDF files there are fewer worries. The body of work found in the PDF file is the substance of the dissertation or thesis, and the inaccessibility of the supplemental supporting media files is unfortunate but may not be detrimental to the main work itself.

### 4.5.2    Replication

The Internet2 Distributed Storage Initiative (Beck and Moore, 1998) and LOCKSS (Lots Of Copies Keeps Stuff Safe – see Reich and Rosenthal, 2001) are projects that NDLTD is working with in order to transparently mirror ETDs and their metadata for faster and more reliable access. It is hoped that such efforts will have multiple benefits. First, there should be faster access if copies of collections are available in all regions of the world, since network delays should be less. Second, loss due to a university having problems with its collection or servers should be minimized. Finally, loss due to deliberate attempts to corrupt an ETD should be minimized, since it will be much harder to make unauthorized changes to an unknown number of widely distributed copies.

## 5    Future Plans

### 5.1    NDLTD growth and organization

NDLTD membership continues to expand, and as time moves forward, more and more of the members shift from pilot to production activities. Since ETD'2003 is being held in Berlin, Germany, there should be growth over the next several years in Eastern Europe in particular. Thanks to support by UNESCO, the "ETD Guide" has been prepared in English, French, and Spanish (Moxley et al., 2002). UNESCO also is supporting training efforts in Latin America (along with ISTEC and OAS) and Africa. It is expected that expansion of NDLTD in these regions will lead to much greater interest in the research in these parts of the world than has occurred in the past.

As NDLTD expands, it is important that its organization evolves. In 2003 it will incorporate in USA as a non-profit (501 c 3) institution with worldwide scope. An international board will complement the existing steering committee to ensure that activities represent the needs of all the regions of the world. Initially, Virginia Tech will serve as the secretariat for NDTLD, providing core services, free of charge.

### 5.2    Enhanced Services

As NDLTD grows, interested parties will continue to contribute to the overall success. Virginia Tech will continue to provide various services, and will continue to seek sponsored funding to expand those services further.  In addition, other key partners are providing core services.

First, VTLS runs a computer to support access to the Union Catalog. Their Virtua software supports multilingual searching and browsing, as well as other standard digital library functionality.

Second, OCLC will expand the Union Catalog.  They will work closely with individual universities, consortia, and other bodies to expand the list of sites from which they can harvest metadata. In addition, they are extracting from WorldCat all the metadata for theses and dissertations, which number roughly 4.5 million. From among those, they will identify the subset that has a URL or other indicator that an ETD is present.  Both the "TD" and the "ETD" collections will be available using the Open Archives Initiative Protocol for Metadata Harvesting.

In addition to the work of these partners, Virginia Tech and others will work to improve and expand services. Based on a study of what might be most helpful, the following list has been developed:

- The ETD-MS metadata format has support for authority control information and, if these fields are populated by ETD archives, they can be used to correctly attribute works to unique individuals and organizations.

- Citations can be extracted from the source ETDs, or from XML representations where such exist, to cross-link the documents in a manner similar to ResearchIndex/CiteSeer (Giles, et al., 1998).

- Recommendations can be generated for individual users or user communities to suggest relevant documents based on the use of the system by other users. Profile-based filtering is an alternative that stores research interests of individual users and then uses these to check for relevance among newly submitted ETDs.

- Currently, the ETD-db software (Atkins, 2002) manages the workflow of ETD submission but it does not handle the process of review and acceptance by the school. This software can be extended or supplemented to fully automate the process.

- Full-text searching can be supported in addition to current access to the Union Catalog that only considers metadata.

- Search facilities can be further enhanced by adding support for cross-lingual searching, where the search terms are provided in one language but the documents are written in another.

In conclusion, we note that with very little funding, a large international digital library has emerged, grown, and improved. This may represent the first of many efforts wherein moving toward digital library technology leads to so many benefits and savings that the digital library is self-sustaining, and expands to provide valuable services worldwide.

## 6    Acknowledgements

## 7    References

Atkins, A. (2002), *Resources for Developers of ETD databases*. Website http://scholar.lib.vt.edu/ETD-db/developer/

Atkins, A., Fox, E. A., France, K. R., and Suleman, H. (2001), *ETD-ms: an Interoperability Metadata Standard for Electronic Theses and Dissertations version 1.00,* NDLTD. Available http://www.ndltd.org/standards/metadata/

Beck, M., and Moore, T. (1998), 'The I-2 DSI Project: An Architecture for Internet Content Channels', in *Computer Networking and ISDN Systems*, **30**(22-23), November, 2141-2148.

Borgman, C. L. (1999), 'What are digital libraries? Competing visions', in *Information Processing and Management*, **35**(3), January, 227-243.

Calado, P., da Silva, A. S., Ribeiro-Neto, B., Laender, A. H. F., Lage, J. P., Reis, D. C., Roberto, P. A., Vieira, M. V., Gonçalves, M. A., and Fox E. A. 'From the Web to a Digital Library: the Web-DL Architecture', JCDL'2003, Third Joint ACM / IEEE-CS Joint Conference on Digital Libraries, 27-31 May, 2003, Houston, to appear.

Dalton, J. and Seamans, N. H. (2003), 'Electronic Theses & Dissertations:Two Surveys of Editors & Publishers**'**, in Fox, E. A., Feizbadi, S., Moxley, J. M. and Weisser, C. R.(eds.), *The ETD Sourcebook: Theses and Dissertations in the Electronic Age*, New York: Marcel Dekker, in press.

Fox, E. A. and Marchionini, G. (1998), 'Toward a Worldwide Digital Library', Guest Editors' Introduction to special section (pp. 28-98) on *Digital Libraries: Global Scope, Unlimited Access. Communications of the ACM,* **41**(4), April, 28-32. http://purl.lib.vt.edu/dlib/pubs/CACM199804

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., Solvberg, I., 'Digital Libraries: A Generic Classification and Evaluation Scheme', in Proceedings European Conference on Digital Libraries, 2001,Springer, Berlin.

Giles, C. L., Bollacker, K. D., and Lawrence, S., 'CiteSeer: An Automatic Citation Indexing System', in Proceedings of Third ACM Conference on Digital Libraries, Pittsburgh pp. 89-98, PA, USA, 23-26 June, 1998.

Gonçalves, M. A., Fox, E. A., Watson, L. T. and Kipp, N. A. (2003) 'Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries', Technical Report TR-03-04, Computer Science, Virginia Tech. http://eprints.cs.vt.edu:8000/archive/00000646/

Lagoze, C., Van de Sompel H., Nelson, M., and Simeon W. (2002), *The Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0*, Open Archives Initiative, June 2002. Available http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

McMillan, G. (2001). 'Do ETDs Deter Publishers? coverage from the 4th Inernational Symposium on ETDs', *College and Research Libraries News,* **62**(6), June, 620-621.

McMillan, G., 'What to Expect from ETDs: Library Issues and Responsibilities.' 4th International Conference on ETDs, 24 March, 2001, Cal Tech.

Moxley, J. M., et al. (2002). *The ETD Guide.* www.etdguide.org

Reich, V., and Rosenthal D. S. H. (2001), 'LOCKSS: A Permanent Web Publishing and Access System', in *D-Lib Magazine*, **7**(6), June, Available http://www.dlib.org/dlib/june01/reich/06reich.html

Saracevic, T. (2000). 'Digital Library Evaluation: Toward Evolution of Concepts', *Library Trends*, **49**(2), 350-369.

Suleman, H. (2002), *ODL-based ETD Union Catalog*. Website http://purl.org/net/etdunion

Urs, S. R., Raghavan, K. S. (2001). 'Vidyanidhi: Indian digital library of electronic theses' In *Communications of the ACM*, **44**(5), May, 88-89.

VTLS (2002), *VTLS*. Website http://www.vtls.com