
ARTICLES

D-Lib Magazine
September 2001

Volume 7 Number 9

ISSN 1082-9873

Networked Digital Library of Theses and Dissertations**Bridging the Gaps for Global Access - Part 1: Mission and Progress**

[Hussein Suleman](#), [Anthony Atkins](#), [Marcos A. Gonçalves](#), [Robert K. France](#), [Edward A. Fox](#)
([hussein](#), [anthony.atkins](#), [mgoncalv](#), [france](#), [fox](#)) @vt.edu
Virginia Tech

[Vinod Chachra](#), [Murray Crowder](#)
([chachra](#), [crowderm](#)) @vtls.com
VTLS Inc.

[Jeff Young](#)
[jyoung@oclc.org](#)
OCLC

Abstract

The Networked Digital Library of Theses and Dissertations (NDLTD) is a collaborative effort of universities around the world to promote creating, archiving, distributing and accessing Electronic Theses and Dissertations (ETDs). Since its inception in 1996, over a hundred universities have joined the initiative, underscoring the importance institutions place on training their graduates in the emerging forms of digital publishing and information access. The outreach and training mission of NDLTD is an ongoing project so in this article we report on the current status of membership and support activities. Recent research has focused on creating a union database that will provide a means to search and retrieve ETDs from the combined collections of NDLTD member institutions. The Virtua system developed by VTLS will serve as the heart of this union database. In order to bridge the gap between the existing distributed institutional archives and a unified collection of ETDs, we have developed a metadata standard especially suited to ETDs - this is then used by partner sites to export their freely-available metadata using the Metadata Harvesting Protocol of the Open Archives Initiative. We also link name authority information into the metadata records to support unique identification of authors and others associated with the works. Additional research efforts include advanced search mechanisms, semantic interoperability, the design and development of multi- and cross-lingual search systems, and software modules that support the development of higher-level services to aid researchers in seeking relevant ETDs.

Introduction

The Networked Digital Library of Theses and Dissertations (NDLTD, see <http://www.ndltd.org>) has emerged as a result of the efforts of thousands of students, faculty, and staff at hundreds of universities around the world, as well as the assistance of interested parties at scores of companies, government agencies, and other organizations. This federation has multiple objectives, including:

- to improve graduate education by allowing students to produce electronic documents, use digital libraries, and understand issues in publishing;
- to increase the availability of student research for scholars and to preserve it electronically;
- to lower the cost of submitting and handling theses and dissertations;
- to empower students to convey a richer message through the use of multimedia and hypermedia technologies;
- to empower universities to unlock their information resources; and
- to advance digital library technology.

Work toward those objectives has proceeded since November 1987, the date of the first meeting devoted to exploring how advanced electronic publishing technologies could be applied to the preparation of electronic theses and dissertations (ETDs). Early efforts are summarized in two D-Lib articles in 1996 and 1997 [[Fox et al., 1996](#); [Fox et al., 1997](#)]. A third article summarizes the first attempts to support, through federated search, access to the collection (see also <http://www.theses.org>) of ETDs that is emerging in distributed fashion [[Powell & Fox, 1998](#)].

NDLTD activities are coordinated by an international steering committee that meets each spring and fall. Its members include those who lead the diverse regional and national efforts that promote efforts regarding ETDs. Committees help with strategic planning, standards (see <http://www.ndltd.org/standards>), training, and meetings. A good deal of effort by steering committee members has gone into fund-raising, so that single and groups of institutions could implement ETD initiatives. There have been national projects in the USA [[Kipp, et al., 1999](#)], South Africa, Germany, Australia, and other countries. Supporting research work has been funded by NSF (in projects IIS-9986089 [[Fox, 2000](#)], IIS-0086227 [[Fox, et al. 2000](#)], IIS-0080748 [[Fox, et al., 2001](#)]), as well as DFG (in Germany) and CONACyT (in Mexico).

At the grass roots level, one line of support for NDLTD emerged from efforts at Virginia Tech, which has developed training materials and workflow management software that have been adapted by diverse groups. Many other projects and programs interested in ETDs have arisen around the world, some independently, but all are welcome to collaborate through the growing federation that is NDLTD. This is important since open sharing of methods helps others know how to address problems as well as ongoing changes in technology. Already there have been four international symposia on ETDs, with approximately 200 attendees at each of the last two. The next two will be held May 30 - June 1, 2002 at Brigham Young University in Provo, Utah, and late spring, 2003, in Berlin. The NDLTD steering committee has its spring meetings in conjunction with the ETD conferences.

These efforts should have a strong positive effect on expanding awareness at universities around the globe. One important agent promoting learning in this arena is the UNESCO International Guide for the Creation of ETDs (see <http://etdguide.org/>). To be released late in 2001 in a number of different languages, this book / web site should help students, faculty, and administrators participate in NDLTD. This should extend the considerable progress already made, as is discussed in the next section.

NDLTD Progress

NDLTD has experienced constant progress since its formation. We have registered growth in all major facets, including membership (with an increasing international participation), collection size, access, multimedia use, and worldwide availability.

Membership

Table 1 shows NDLTD membership as of August 2001. In less than two and a half years, NDLTD has more than doubled the number of registered members (from 59 members in May 1999). There are currently 120 members; 52 U.S. universities, 52 non-U.S. universities, and 16 institutions, regional centers and organizations (such as UNESCO). These various partners represent 23 countries: Australia, Brazil, Canada, China, Colombia, Germany, Greece, Hong Kong, India, Italy, Mexico, Netherlands, Norway, Russia, Singapore, South Africa, South Korea, Spain, Sudan, Sweden, Taiwan, the USA, and the United Kingdom. These numbers also emphasize the growth of global interest in NDLTD as international participation grew from less than one third in 1998 to half of the total membership in 2001. Also, by early 2002, at least 11 of the registered NDLTD members will have started requiring mandatory submission of electronic theses and dissertations. (In the table below, those are marked with an asterisk.)

USA Universities (524)	International Universities (512)	Institutions (156)
Air University (Alabama)	Alicante University (Spain)	Cinemedia
Alicante University	Australian National University (Australia)	Coalition for Networked Information
Baylor University	Biblioteca de Catalunya (Spain)	Committee on Institutional Cooperation
Brigham Young University	Chinese University of Hong Kong (Hong Kong)	Consorti de Biblioteques Univers. Catalunya
California Institute of Technology	Chungnam National U., Dept of CS (S. Korea)	Diplomica.com
Clemson University	City University, London (UK)	Dissertationene Online
College of William and Mary	Curtin University of Technology (Australia)	Dissertation.com
Concordia University (Illinois)	Darmstadt University of Technology (Germany)	ETDweb
East Carolina University	Freie Universitat Berlin (Germany)	Ibero-American Sci. & Tech. Ed. Cons. (ISTEC)
East Tennessee State University*	Gerhard Mercator Universitat Duisburg (Germany)	National Documentation Centre (NDC, Greece)
Florida Institute of Technology	Griffith University (Australia)	National Library of Portugal
Florida International University	Gyeongsang National University, Chinju (Korea)	OhioLINK
George Washington University	Humboldt-Universität zu Berlin (Germany)	OCLC
Louisiana State University*	Indian Institute of Technology, Bombay (India)	Organization of American States (OAS)
Marshall University	Lund University (Sweden)	SOLINET
Massachusetts Institute of Technology	McGill University (Canada)	Sudanese National Electronic Library (Sudan)
Miami University of Ohio	National Sun Yat-Sen University (Taiwan)	Solinet UNESCO
Michigan Tech	Nanyang Technological University (Singapore)	
Mississippi State University		
Montana State University		
Naval Postgraduate School		
New Jersey Institute of Technology		
New Mexico Tech		
North Carolina State University*		

North Western University	National University of Singapore (Singapore)
Pennsylvania State University	Rand Afrikaans University (South Africa)
Regis University	Rhodes University (South Africa)*
Rochester Institute of Technology	Shanghai Jiao Tong University (China)
Texas A&M University	St. Petersburg State Technical U. (Russia)
University of Colorado	State University of Campinas (Brazil)
University of Florida	Sudanese National Electronic Library (Sudan)
University of Georgia	Universidad de las Américas Puebla (Mèxico)
University of Hawaii at Manoa	Universitat Autònoma de Barcelona (Spain)*
University of Iowa	Universitat d'Alacant (Spain)
University of Kentucky	Universitat de Barcelona (Spain)
University of Maine*	Universitat de Girona (Spain)
University of North Texas*	Universitat de Lleida (Spain)
University of Oklahoma	Universitat Oberta de Catalunya (Spain)
University of Pittsburgh	Universitat Politècnica de Catalunya (Spain)
University of Rochester	Universitat Politècnica de València (Spain)
University of South Florida	Universitat Pompeu Fabra (Spain)
University of Tennessee, Knoxville	Universitat Rovira i Virgili (Spain)
University of Tennessee, Memphis	Université Laval (Québec, Canada)
University of Texas at Austin*	University of Bergen (Norway)
University of Utrecht	University of Antioquia (Medellin, Colombia)
University of Virginia	University of British Columbia (Canada)
University of West Florida	University of Guelph (Ontario, Canada)
University of Wisconsin, Madison	University of Hong Kong*
Vanderbilt University	University of Melbourne (Australia)
Virginia Commonwealth University	University of Mysore (India)
Virginia Tech*	University of New South Wales (Australia)
West Virginia University*	University of Pisa (Italy)
Western Michigan University	
Worcester Polytechnic Institute	

University of Queensland (Australia)
University of Sao Paulo (Brazil)
University of Sydney (Australia)
University of Utrecht (Netherlands)
University of Waterloo (Canada)
Uppsala University (Sweden)
Wilfrid Laurier University (Canada)

Table 1. NDLTD Membership

Collection Size

The number of ETDs across the NDLTD universities/institutions has grown at an even faster pace. From a few dozen at Virginia Tech in 1996, to 4,328 ETDs at 21 institutions in March 2000, we accounted for a total of 7,268 ETDs at 25 member institutions in July 2001. Table 2 shows a breakdown of the current numbers of ETDs as of July 2001 organized by member institution. This data is largely the result of an on-line survey conducted by Gail McMillan and represents only those institutions that responded to the survey.

University/Institution	ETD Collection size
ADT: Australian Digital Thesis Program (Australia)	238
University of Bergen (Norway)	45
California Institute of Technology	2
Consorci de Biblioteques Universitaries de Catalunya (Spain)	151
East Tennessee State University	106
Humboldt-University (Germany)	430
Louisiana State University	3
Mississippi State University	33
MIT	62

North Carolina State University	301
Pennsylvania State University	83
Pontifical Catholic University (PUC) (Brazil)	90
Gerhard Mercator Universitat Duisburg (Germany)	126
Universitat Politecnica de Valencia (Spain)	189
University of Florida	174
University of Georgia	121
University of Iowa	6
University of Kentucky	19
University of Maine	27
University of North Texas	337
University of South Florida	25
University of Tennessee	12
University of Tennessee, Knoxville	28
Uppsala University (Sweden)	178
Virginia Tech	3393
West Virginia University	1006
Worcester Polytechnic Institute	83
TOTAL	7268

Table 2. ND LTD collection size

These statistics do not take into account scanned theses and dissertations, which make up a substantial portion of the total NDLTD collection. There are 26 scanned documents at the New Jersey Institute of Technology, 150 at the University of South Florida, 5,581 at MIT, and 12,000 at the National Documentation Center in Greece. These result in a total of 17,763 scanned theses and dissertations at these institutions, and quite conceivably thousands of unreported ones at other institutions.

Access Statistics

To demonstrate the potential of NDLTD for global access and sharing of the knowledge produced by universities worldwide, we have periodically analyzed the access logs of the Virginia Tech ETD (VT-ETD) collections. The results for the period 1997-2000 are shown below in Table 3.

	1997/98	1998/99	increase 1997/98- 1998/99	1999/00	increase 1998/99- 1999/00
Requests for PDF files (mostly full ETDs)	221,679	481,038	117.0%	578,152	20.2%
Requests for HTML files (mostly tables of contents and abstracts)	165,710	215,539	30.1%	260,699	21.0%
Requests for multimedia	1,714	4,468	160.7%	12,633	182.7%
Distinct files requested	6,419	21,451	234.2%	16,409	-23.5%
Distinct hosts served	29,816	57,901	94.2%	87,804	51.6%
Average data transferred daily	156,089 KMBB	219,132 KMBB	40.4%	382 MbMB	74.4%
Data transferred	55,637 MbGB	78,107 MbGB	40.4%	137 GbGB	75.6%

Table 3. Access Log Statistics from the VT-ETD collection

We can see that the number of accesses tends to increase each year. As the collection grows and gains popularity, the number of accesses will most likely continue to increase.

More specifically, Table 4 indicates that each of the seven countries with the most accesses has an increasing number of accesses each year (with the exception of Germany in the 97/98 - 98/99 period). The United Kingdom, and surprisingly Malaysia, dominated accesses from outside the US. The other accessing countries are all European, a fact that is probably related to advances in network infrastructure in those countries.

International Domain	1997/98	1997/98 rank	1998/99	1998/99 rank	Increase 1997/98- 1998/99	1999/00	1999/00 rank	Increase 1998/99- 1999/00
United Kingdom	6,735	1	11,347	1	68.5%	25,583	1	125.5%
Malaysia	876	16	4,190	6	378.3%	16,147	2	285.4%
France	2,138	7	4,797	5	124.4%	14,960	3	211.9%
Germany	6,727	2	3,374	9	-49.8%	14,384	4	326.3%
Canada	3,413	4	9,632	3	182.2%	13,543	5	40.6%
Spain	590	18	3,647	8	518.1%	9,918	6	171.9%
Italy	1,430	12	3,095	10	116.4%	9,300	7	200.5%

Table 4. Access by Non-US Sites

Multimedia Use in ETDs

One of the main objectives of NDLTD is to promote student creativity through the use of diverse types of multimedia content in ETDs, while making students comfortable with the use of this technology to exploit richer modes of self-expression.

Table 5 indicates how much of this objective has been achieved in the VT-ETD collection, with a breakdown of the 8,056 multimedia files contained in a selection of 2,180 available ETDs. This illustrates both that authors are beginning to shift towards non-textual media and that some are moving away from the early single-file paradigm of digitization.

File type	Examples	Count
Still image	BMP, DXF, GIF, JPG, TIFF	328
Video	AVI, MOV, MPG, QT	58
Audio	AIFF, WAV	18
Text	PDF, HTML, TXT, DOC, XLS	7601
Other	Macromedia, SGML, XML	51

Table 5. Multimedia use in VT-ETD collection

Worldwide Release

In terms of copyright, a significant issue is whether to allow the electronic document to be viewed worldwide, on campus only, or not at all. The “mixed” case, which is a unique capability of electronic documents, occurs when some portions (e.g., particular chapters or multimedia files) have restricted access while others are more widely available. The majority of Virginia Tech students allow their documents to be viewable worldwide (see Figure 1) - but some initially choose not to grant worldwide access in order to protect their publication rights. To address this concern, there are ongoing discussions with publishers to help them understand the goals and benefits of NDLTD [[NDLTD, 1999](#)]. We are pleased to see a change in attitude by some publishers over the course of the project. The American Chemical Society developed a policy more favorable to NDLTD as a result of lengthy discussions and the American Physics Society has been receptive to issues concerning the Open Archives Initiative and NDLTD.

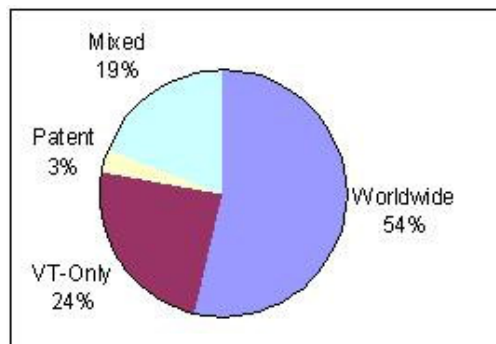


Figure 1. Student and committee choice for ETD availability from Virginia Tech (2668 ETDs as of July 17, 2000).

Standards Activity

In order to support many of the current and future research and service-related activities, work has begun to define standards that will enable more consistent exchange of information in an interoperable environment. Among the first of these projects is ETDMS - the Electronic Thesis and Dissertation Metadata Standard - and a related project for name authority control.

Electronic Thesis and Dissertation Metadata Standard (ETDMS)

ETDMS was developed in conjunction with the NDLTD, and has been refined over the course of the last year. The initial goal was to develop a single standard XML DTD for encoding the full text of an ETD. Among other things, an ETD encoded in XML could include rich metadata about the author and work that could easily be extracted for use in union databases and the like. During initial discussions it became clear that the methods used by different institutions to prepare and deal with theses and dissertations would make it all but impossible to agree on a single DTD for encoding the full text of an ETD. Many institutions were unwilling or unprepared to use XML to encode ETDs at all.

Thus, instead of an XML DTD for encoding the full text of an ETD, ETDMS emerged as a flexible set of guidelines for encoding and sharing very basic metadata regarding ETDs among institutions. Separate work continues in parallel on a suite of DTDs, building on a common framework, for full ETDs.

ETDMS is based on the Dublin Core Element Set [[DCMI, 1999](#)], but includes an additional element specific to metadata regarding theses and dissertations. Despite its name, ETDMS is designed to deal with metadata associated with both paper and electronic theses and dissertations. It also is designed to handle metadata in many languages, including metadata regarding a single work that has been recorded in different languages. The ETDMS standard

[[Atkins, et al., 2001](#)] provides detailed guidelines on mapping information about an ETD to metadata elements.

ETDMS already is supported as an output format for the Open Archives interface to the Virginia Tech ETD collection. ETDMS will be accepted as an input format for the union catalog currently being developed in conjunction with VTLS [[VTLS, 2001](#)]. NDLTD strongly encourages use of ETDMS.

Authority Linking

Each reference to an individual or institution in an ETDMS field should contain a string representing the name of the individual or institution as it appears in the work. In addition, these references also may contain a URI that points to an authoritative record for that individual or institution. Associating authority control with NDLTD seems particularly appropriate since universities know a great deal about those to whom they award degrees and since a thesis or dissertation often is the first significant publication of a student.

The “NDLTD: Authority Linking Proposal” [[Young, 2001](#)] identifies several goals for a Linked Authority File (LAF) system to support this requirement:

- LAF records should be freely created and shared among participants. While a central authority database is an option, the LAF design expects the database to be distributed to share cost. Individual participants or groups should be able to host a copy of the LAF database and share changes they make to local copies of LAF records with other hosts using the Open Archives Initiative (OAI) protocol [[Lagoze and Van de Sompel, 2001](#)]. The mechanism for keeping records in sync is described in the proposal.
- The URIs should be meaningful and useful to anyone outside NDLTD’s domain. A benefit of using the OAI protocol is that individual LAF records will be accessible via an OAI GetRecord request (discussed in the second part of this article).
- The URIs should be persistent and current. This raises a number of challenges, such as duplicate resolution. By using PURLs [[OCLC, 2001](#)] in ETDMS records, the underlying OAI GetRecord URLs can be rearranged without affecting the ETDMS records that rely on them.
- The model should be scalable and applicable beyond NDLTD. The LAF model was designed to work entirely with open standards and open-source software.

The LAF design has other advantages over alternatives such as the Library of Congress Name Authority Database [[Library of Congress, 2001](#)]. Only the level of participation among decentralized participants limits the coverage of the collection. Because the records are based on XML, the content of LAF records can be as broad or narrow as needed. Finally, because they are distributed using the OAI protocol, multiple metadata formats can be supported.

Future of NDLTD

The statistics presented illustrate that the production and archiving of electronic theses and dissertations is fast becoming an accepted part of the normal operation of universities in the new electronic age. NDLTD is dedicated to supporting this trend with tools, standards, and services that empower individual institutions to set up and maintain their own collections of ETDs. At the same time NDLTD promotes the use of these ETDs through institutional websites as well as portal-type websites that aggregate the individual sites and create seamless views of the NDLTD collection.

Ongoing research and service-provision projects are addressing the problems of how to merge together the currently distributed and somewhat isolated collections hosted at each member institution. The second part of this article discusses some of these projects in detail, including development of the Union Catalog Portal based on VTLS’s Virtua system and the myriad of

research efforts investigating how to provide better services to researchers with specific information-seeking needs and behaviors.

References

Atkins, Anthony, Edward A. Fox, Robert France and Hussein Suleman (editors). 2001. *ETD-ms: an Interoperability Metadata Standard for Electronic Theses and Dissertations -- version 1.00*. Available from <<http://www.ndltd.org/standards/metadata/ETD-ms-v1.00.html>>.

DCMI. 1999. *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. Available from <<http://www.dublincore.org/documents/dces/>>.

Fox, Edward A. 2000. Core Research for the Networked University Digital Library (NUDL), NSF IIS-9986089 (SGER), 5/15/2000 - 3/1/2002. Project director, E. Fox.

Fox, Edward A., John L. Eaton, Gail McMillan, Neill A. Kipp, Laura Weiss, Emilio Arce, and Scott Guyer. 1996. National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources, *D-Lib Magazine*, September 1996. Available at <<http://www.dlib.org/dlib/september96/theses/09fox.html>>.

Fox, Edward A., Brian DeVane, John L. Eaton, Neill A. Kipp, Paul Mather, Tim McGonigle, Gail McMillan, and William Schweiker. 1997. Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources, *D-Lib Magazine*, September 1997. Available at <<http://www.dlib.org/dlib/september97/theses/09fox.html>>.

Fox, Edward A., Royca Zia, and Eberhard Hilf. 2000. Open Archives: Distributed services for physicists and graduate students (OAD), NSF IIS-0086227, 9/1/2000-8/31/2003. Project director, E. Fox (w. Royce Zia, Physics, VT, and E. Hilf, U. Oldenburg, PI on matching German DFG project).

Fox, Edward A., J. Alfredo Sánchez, and David Garza-Salazar. 2001. High Performance Interoperable Digital Libraries in the Open Archives Initiative, NSF IIS-0080748, 3/1/2001-2/28/2003. Project director, E. Fox (with co-PIs J. Alfredo Sánchez, Universidad de las Américas-Puebla --- UDLA, and David Garza-Salazar, Monterrey Technology Institute --- ITESM, both funded by CONACyT in Mexico).

Kipp, Neill, Edward A. Fox, Gail McMillan, and John L. Eaton. 1999. *FIPSE Final Report*, 11/30/99. Available from <<http://www.ndltd.org/pubs/FIPSEfr.pdf>> (PDF version) and <<http://www.ndltd.org/pubs/FIPSEfr.doc>> (MS-Word version).

Lagoze, Carl and Herbert Van de Sompel. 2001. *The Open Archives Initiative Protocol for Metadata Harvesting*. Open Archives Initiative. January 2001. Available from <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>.

Library of Congress. 2001. Program for Cooperative Cataloguing Name Authority Component Home Page. Available from <<http://www.loc.gov/catdir/pcc/naco.html>>.

NDLTD. 1999. *Publishers and the NDLTD*. NDLTD, July 1999. Available from <<http://www.ndltd.org/pubshrs/>>.

OCLC. 2001. Persistent URL Home Page. Dublin, OH: OCLC Online Computer Library Center. Available from <<http://purl.oclc.org/>>.

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

VTLS. 2001. Virtua ILS. Available from <<http://www.vtls.com/products/virtua>>.

Young, Jeffrey A. 2001. NDLTD: Authority Linking Proposal. Dublin, OH: OCLC Online

Computer Library Center. Available from <<http://alcme.oclc.org/ndltd/AuthLink.html>>.

Copyright 2001 Hussein Suleman, Anthony Atkins, Marcos A. Gonçalves, Robert K. France, Edward A. Fox, Vinod Chachra, Murray Crowder, and Jeff Young

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)
[Previous Article](#) | [Next Article](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

[DOI: 10.1045/september2001-suleman-pt1](#)