

Networked Digital Library of Theses and Dissertations

Edward A. Fox
Department of Computer Science
Virginia Tech, Blacksburg, VA 24061 USA

Abstract

The Networked Digital Library of Theses and Dissertations aims to enhance learning by enabling authors of these documents to create them electronically and upload them directly into a worldwide federated digital library, where suitable access controls are enforced. This project can be viewed as representative of digital library efforts, and is easily understood by employing the “5S Framework” that has been developed to describe information systems. Scenarios for the key user groups illustrate how the project operates, providing a foundation for collaboration among universities - supported by innovative technology - that addresses a number of important research problems.

Keywords: Digital libraries, dissertations, graduate education, theses

1. Introduction

As of June 1999, the Networked Digital Library of Theses and Dissertations, NDLTD [1-4], includes 62 universities and other interested parties working to enhance education and research. Universities from around the globe, along with a number of supporting institutions (e.g., the Coalition for Networked Information and UNESCO), encourage students to prepare theses (undergraduate or graduate) and dissertations electronically, submit them into a digital library [5], and facilitate access by other students and researchers. It is hoped that this federated initiative will spread to every country and university, ensuring that the next generation of scholars is prepared to create expressive electronic documents, and able to work with worldwide digital libraries [6].

1.1 Digital Library Case Study

NDLTD encompasses all of the key aspects of digital library efforts. It is a project, with a home page (<http://www.ndltd.org>), and a governing body (i.e., a Steering Committee) that meets in Washington, D.C. twice annually, near the middle of April and near the middle of September. It includes as its collection all of the bibliographic records, metadata, documents, and related works made available by its members or other supporters. The focus of the collection is on theses and dissertations, but reports and other documents are also welcome; NDLTD may gradually broaden to include a richer diversity of content as it evolves into a Networked University Digital Library (NUDL). The collection can be accessed through browsing and federated searching mechanisms from <http://www.theses.org> as well as through a variety of connections to repositories run by individual members or groups of members. Students wishing to learn how to create electronic theses and dissertations (ETDs [7, 8]) can benefit from a training site rich in tutorials and multimedia explanations (<http://etd.vt.edu>). They are encouraged to use standards (e.g., SGML, XML, PDF, MPEG, JPEG) in order to facilitate preservation.

1.2 View from the May 1999 Workshop

Also associated with NDLTD is an annual workshop. The first occurred in the summer of 1998 in Memphis (TN) while the second was in May 1999 at Virginia Tech (Blacksburg, VA), with seventy attendees. A workshop committee is reviewing several proposals for the 2000 event as it proceeds to select another suitable venue.

Some of the attendees had been involved in NDLTD since early 1996. Then, funding by the Southeastern University Research Association (SURA) allowed regional expansion, which slightly preceded the U.S. Department of Education support of a national project in September 1996. Also attending the meeting were representatives of several Canadian universities, which began joining in 1997, reflecting the progress that moved the initiative from a national to an international enterprise (with national efforts beginning that year in Australia and Germany).

At the workshop members decided to carry forward the project through a series of committees, each charged with addressing key concerns of the group:

- workshop planning (to select a venue and assemble a conference committee),
- training (to ensure that materials are prepared in different languages for students and trainers),
- standards (to identify and document standards facilitating preservation, that are easily usable),
- publisher relations (to ensure that publishers understand and support the initiative),
- software (to enhance automation, browsing, searching, and information exchange), and
- statistics and reporting (to document progress and determine effects on learning and research).

Attendees shared their experiences and solutions. Many learned of new approaches to difficult problems they had faced, and all left encouraged to push forward at their home institution. Some decided to join NDLTD, and begin pilot efforts. A number decided to shift from a pilot effort to allow all interested students to submit their theses electronically. Others decided to set a date whence all students would be required to submit an electronic thesis or dissertation (ETD). Though these successive stages demonstrate increasing levels of commitment, they simplify the situation accordingly, since it is easier to handle all theses electronically than to have all or most submissions on paper. Virginia Tech and West Virginia University, the first two institutions to require ETDs, explained the smooth procedures in place. Statistics on the rapidly growing number of accesses to the Virginia Tech collections indicated the strong demand for ETDs from around the world, encouraging other universities to make available to their own students this vehicle for disseminating scholarly findings. Finally, attending universities agreed to work together to ensure the continuing expansion of NDLTD and the enhancement of the services it provides.

1.3 5S Framework

At the heart of NDLTD is the aim of enhancing graduate education through the application of electronic publishing and digital library technologies. Accordingly, we have developed a descriptive framework encompassing these technologies — 5S — referring to Societies, Scenarios, Spaces, Structures, and Streams. We argue that this framework not only helps with the design and development of digital libraries, but also should help students more easily understand new related technologies.

5S is particularly oriented to describe information systems. We view digital libraries as high-end or “super” information systems that integrate a wide variety of more specialized technologies, and so benefit strongly from such a powerful framework. Advanced information systems often involve multimedia information and distributed processing, both of which require special support for Streams. Various approaches to information organization - whether in collections, using databases, supported by indices as in geographic information systems, through graphs (as in hypertext), or as complex objects - involve suitable Structures. Scientific visualization, virtual reality simulations, vector space or probabilistic or conceptual searching, and 2D or 3D graphics interfaces all make use of Spaces. Since digital libraries provide a range of services, supporting various types of information needs in tailored fashion, and are often designed based on story-like descriptions of interactions, the use of Scenarios is particularly important. Finally, since digital libraries are built to serve particular target users or user communities, it is essential that the Societies involved be carefully studied (i.e., we extend our earlier 4S model [9] by adding societies).

Building on this framework, we discuss other aspects of NDLTD in this paper. The next section explains more about NDLTD by describing representative Scenarios for various Societies involved. Section 3 broadens the focus to consider collaboration at the international level. Section 4 details some of the underlying technology developed. Section 5 explains challenges still faced, while Section 6 concludes the paper.

2. Scenarios

As explained in Section 1.3, digital library services can be explained through Scenarios for each of the Societies involved. We discuss the most important scenarios in the following subsections. Most are feasible or could be supported with a moderate amount of work.

2.1 Scenarios for Users

The largest potential community related to NDLTD is that of users of the growing collection. During 1998, about thirty-seven thousand different sites accessed Virginia Tech ETDs. Once a really large and comprehensive collection emerges as the result of scores of universities each contributing hundreds or thousands of works per year, we expect that the base of users will number in the millions. This will include the hundreds of thousands of students engaged in graduate work and the millions of scholars engaged in research investigations around the globe.

A graduate student user is likely to wish to find works to guide personal research. The user may have a topic well in mind, know how to search, and wish to make sure that the problem selected has not already been solved by another. This situation requires a comprehensive search, somewhat similar to those called for in legal cases. In such situations the cost of missing a relevant document can be extraordinarily high (e.g., turning years of work into a spurious effort).

Another scenario involves a graduate student hoping to identify a topic for research. One search goal may be a suitable set of highly related studies, where the detailed literature reviews and bibliographies included can serve as an introduction to the topical area. After these chapters are read, the corresponding bibliographies may lead directly to other interesting works. Alternatively, newer leads may emerge from a citation database that may be built inside NDLTD (or, perhaps, constructed jointly with the Institute for Scientific Information by extending their indexes).

Building upon the annotation server prototyped by Todd Miller early in 1999, users may decide to add notes in the form of private annotations whenever they read something in NDLTD. These will be stored on their local server, so whenever the user connects to NDLTD, all past annotations become available. Using these, an annotated bibliography could be easily produced, which might become a chapter in one's own ETD.

Research users may desire a number of other specialized services. Focus group discussions recorded by Todd Miller in 1998 suggested several. It would be helpful to have a program to analyze an ETD and extract / generate from it a glossary of important terms. It would be useful to take a small number of literature reviews and compile an overall summary for a sub-field. This might indicate which references were cited repeatedly, and group together the varied comments around each such reference. Somewhat more difficult to develop might be a tool that would summarize the open problems mentioned repeatedly in a small collection of ETDs, to help in the search for good research topics. Many other services and scenarios might be of value for users of NDLTD. Specialized ones might address the needs of particular types of users, e.g., teachers hoping to use parts of a dissertation in a class presentation, or chemists looking for works that employ a particular type of methodology.

2.2 Scenarios for an Author

At the heart of the educational aspect of NDLTD is the notion of students learning by doing. A student may be motivated to learn more about electronic publishing and digital libraries if that learning helps make one's own thesis accessible to a larger number of potential readers. Students may learn more about multimedia technologies if using them allows an ETD to be more expressive, possibly through more extensive use of images, audio, and/or video. A student may learn a great deal about the world of publishing through writing to a publisher that is handling a related journal submission to explain that they want their ETD to simultaneously be made world accessible, as has already been endorsed by such publishers as ACM, IEEE-CS, and Elsevier. A student may understand about preservation after creating an easy to analyze XML version of their ETD, instead of a Word version that may become unusable in a few years.

Most students relate to NDLTD through one or more of three key scenarios. First, during their research, they may be a user of NDLTD (see Section 2.1 above), studying interesting results and perhaps finding some useful bibliographic references to support further investigation. Second, they will use the NDLTD submission software to upload their ETD, thus adding their work and related metadata to NDLTD. Third, they must connect with the rights and permissions aspect of NDLTD, filling in the Approval Form with their faculty committee and specifying terms and conditions of access to their work. Finally, it is hoped that their ETD may lead others to contact them to offer employment, ask for more details, make suggestions on future research, or suggest collaboration.

2.3 Scenarios for Author-Researcher Collaboration

Several teachers in South America have contacted this author about a masters project and report he supervised that gives a detailed tutorial on the AuthorWare software package, suitable for use in courses on multimedia. One student in South Africa was contacted by a researcher in Berkeley to suggest collaboration, based on access to the student's ETD. Study of the access logs of one Virginia Tech chemistry student's ETD indicated that within a few weeks of submission, more than a dozen different groups working on similar types of investigation had downloaded that ETD.

While these cases give specific instances of benefit from ETDs, it may be more useful to describe general scenarios related to collaboration. First, there are cases where one student develops a method or approach that can be applied by others. Second, there are situations where tools or data sets are developed that are appropriate for re-use. Third, some students may develop theories that others may apply or validate through experimental investigations. Fourth, one student's contribution may lead to another student facing the challenge to improve upon prior work, by developing a faster or more efficient solution, possibly at the same time validating earlier findings by replicating them.

Collaborations often are supported by shared artifacts, such as ETDs. With Todd Miller's annotation tool, an author may allow another to attach public annotations to their own ETD, thus making the comments of collaborators to become available whenever the abstract of the ETD is read. In a less public situation, new collaborators may discuss an ETD by sending comments through email or by communicating using phone, facsimile, or letters. If an ETD is represented using PDF, copies of the ETD may be exchanged with notes attached by tools like Adobe's Acrobat software. In any case, having large numbers of ETDs available for essentially no cost should make it possible for scholars in remote regions to study theses that otherwise might have been out of reach, and should make it easier for them to contact the authors, engaging in mentoring or learning activities.

2.4 Scenarios for Supporting Librarians

Librarians play many roles regarding NDLTD. Once a student has added an ETD to a local collection, and it has been suitably approved, one librarian may catalog the work, using information in the ETD as well as metadata provided during the course of submission. That librarian may (semi-automatically) prepare a MARC record for the local catalog. In addition, a record may be created in a local database system to allow searching of the full-text of the ETD using a local search engine (e.g., the OpenText system used at Virginia Tech, or prototype services developed at Virginia Tech to use OCLC's SiteSearch system or IBM's Digital Library).

Another librarian may implement access restrictions requested by the student, so that part or all of the work becomes accessible only to the local campus community. That librarian may change the access situation later, perhaps shifting from campus to world access, as may happen with a chemistry work that relates to a journal article just published by the American Chemical Society (since their copyright policy supports such as change).

A third librarian, charged with handling preservation, may work on the collection of ETDs filed several years ago. First, the librarian may move the documents to a newer computer or set of storage devices, so that online access continues to be supported. Second, if there are SGML files involved, a program developed at Virginia Tech may be run to generate a new collection of HTML files, using the most recent standard version of the HTML specifications, so this rendering of the ETD can benefit from the newest Web technology. Finally, some conversions may be run on multimedia files, so that old standard forms give way to newer, more widely supported versions. In such cases, the original submission may be left, so users can select from the original authored form or a new rendering that is easier to use.

Finally, other librarians may help various ETD users. Reference librarians may help students find interesting ETDs. Others may help students preparing their ETDs, perhaps when using complex devices such as an editing system for digital video. Some librarians may engage in training sessions, such as occurs in the periodic workshops run so students can learn more about the ETD requirement. A smaller number of librarians may support advanced students

who desire specialized training regarding multimedia information or markup languages (e.g., SGML, XML).

2.5 Scenarios for Graduate Education

In some cases, the Graduate School on a campus may run the training workshops instead of, or along with, librarians. There may be a walk-in service for students unfamiliar with procedures, who come (in person, or virtually) to the Graduate School to request assistance.

In many cases, Graduate Schools are responsible for checking theses, to be sure that all campus policies are enforced. In addition, they may want to check for plagiarism, by submitting one ETD and having a special program make sure that it is not highly similar to any other ETD.

Once an ETD is available, for cases where a student wishes to have UMI (University Microfilms International, of Ann Arbor, Michigan) archive their work and include an entry in their *Dissertation Abstracts* database, the Graduate School may then notify UMI. This may involve reporting that:

- a dissertation has been received;
- a URN has been assigned, and is passed on to UMI;
- fees have been paid for UMI services, and will be forwarded as part of a batch payment;
- a copy of the form filled in by the student will be mailed.

Other scenarios apply regarding NDLTD. It is encouraged that each NDLTD member determine what scenarios are appropriate in their situation, what others are desirable, and which ones they might develop into new services to be shared with other NDLTD members.

3. An Opportunity for Worldwide University Collaboration

Many of the scenarios described in Section 2 explain ways in which students and researchers may collaborate regarding their research. Other scenarios indicate how librarians and graduate school staff may work with students and interested parties to add or utilize ETDs. The following subsections extend this discussion and focus it on university-university collaboration.

3.1 Lack of Exposure to Research Abroad

Today, outside of the activities related to NDLTD, there is very little exposure of students and researchers to the graduate research carried out at other institutions. There are four principal mechanisms for such exposure to develop.

First, there is the sharing of research supported by UMI. According to figures reported by UMI's employee William Savage, less than sixty-thousand works are received by them each year. These account for almost all of the dissertations from USA and Canada, as well as almost all of the masters theses from Canada. Figures are not available to this author regarding how many copies of *Dissertation Abstracts* are sold, or how many copies of theses or dissertations are sold by UMI to interested parties from the stock of about 1.5 million works in their archive. An estimated upper bound might be computed on the number of copies sold, though. If we assume that total income is about \$10M per year and that a single copy costs \$50, UMI might sell 200,000 copies per year. That is much less than one sale per thesis or dissertation in their collection, and no more than about 4 accesses per year to each of

the new works submitted over the last year, if all sales were concentrated on those items. Larger numbers of accesses may result from new UMI services, including free viewing of parts of dissertations that since 1997 have been scanned and made available as PDF page images (300 dpi, black and white, captured from microfilm).

Second, there is sharing based on interlibrary loan. Records at Virginia Tech indicate that this occurs at a low level. Circulation records show that in the first six years after a paper thesis was submitted, it circulated about 2 times per year. Dissertations circulated about 3 times per year in the same period. Since most use is local, we might assume that an average thesis or dissertation would be loaned out less than once a year.

Third, regarding access to international research from inside the U.S., there is little real support. The main arrangement for this process is through the agreements made by the Council of Library Resources. Their Chicago facility has about 750,000 dissertations collected from abroad, mostly from European universities. However, these are only available onsite. Further, there is no electronic or even card catalog, and the books are organized on shelves according to size of book, and then alphabetically according to author name. It appears that there is little access feasible under these circumstances.

Fourth, there are various methods for sharing works in less formal fashions. Some authors are contacted by interested parties who desire a copy. In recent years, authors have begun to post their works on Web sites or in departmental repositories. Some countries collect dissertations in a national library or other depository, which can be visited by interested parties. While industrious researchers make use of all of these mechanisms, it seems unlikely that more than say one access per year on average occurs across university boundaries through such mechanisms.

In summary, we note that U.S. dissertations and Canadian theses are made available through UMI, but on average are probably only sold in relatively small numbers each year. Other theses and dissertations have very low circulation, and particularly low re-use across university boundaries.

3.2 Spread of Interest

In contrast to the situations noted in Section 3.1, ETDs at Virginia Tech's site are downloaded hundreds of times each year, on average. The number of sites accessing the collection has increased rapidly since 1996 when the collection first became available. The 1998 figure on distinct IP addresses was about 37,000 and the number of downloads numbered in the hundreds of thousands, for a collection of under 2000 works. On average, this translates into at least a ten-fold, and maybe a hundred-fold increase in accesses relative to other mechanisms. In addition, log analysis regarding the Virginia Tech collection has shown rapid increase in the accesses from various countries, over the last three years.

Interest in NDLTD is also reflected by the activities at universities exposed to the project. At least a few hundred universities have heard of the project. Visits and discussions involving the Virginia Tech team almost certainly bring to over one hundred the number of universities that either are members or have some clear interest in ETDs.

As is the case in other initiatives involving diffusion of innovation, spread of interest occurs in accord with a variety of factors. The Virginia Tech team has tried to manage this process using a number of approaches:

- contacting universities in the southeast of the U.S., supported by early funding from SURA;
- visiting universities that a member of the project team has other business with;
- visiting universities that request a presentation about NDLTD;
- offering to visit universities that have large graduate programs;
- contacting universities that have active initiatives in the digital library area;
- contacting universities that are well-known;
- repeating visits when interest or opportunity makes that appropriate;
- visiting at least one university in each country that a team member travels to;
- assisting groups that are involved in NDLTD upon request;
- encouraging national, regional, or state groups, as well as other consortia, to join;
- co-authoring or endorsing proposals submitted by NDLTD members;
- using videoconferencing whenever possible to reach remote sites;
- offering to give presentations at conferences and workshops that have broad attendance, particularly by members of the Graduate School, Library, or Digital Library communities.

Hundreds of talks have been given. Additional presentations are scheduled. Other members of NDLTD are also engaged in this dissemination process, which we hope will lead to a number of sites becoming leaders in their nation or region.

3.3 Needs for Joint Work

A key area for joint work involves helping with the spread of NDLTD. There are hundreds of universities that can learn about and join NDLTD. Current members can be nurtured to move toward the stage of requiring ETD submission. Efforts in this regard are probably the most important that can be undertaken to support NDLTD.

Other support is possible to provide suitable infrastructure for NDLTD. One aspect involves the Web sites developed at Virginia Tech, which have been adapted at other locations. It would be helpful to streamline this process, so new sites can more quickly come online with their Web pages. In particular, it would help to have translations of the site into major languages for the various countries involved, with suitable support for character sets.

Part of the activity at an NDLTD site involves workshops, online, and one-on-one training and assistance for students who will prepare an ETD. Since students use a broad diversity of software while preparing an ETD, and since new versions of that software arrive continuously, it is important to continually add to and update the training resources developed for NDLTD. This can easily be undertaken in a distributed fashion, and will be coordinated by the new committee being formed to focus on training.

The other committees formed at the May 1999 Workshop also provide points of focus for joint work. Thus there is need to identify suitable standards, assist with training about them, locate tools that facilitate conversion from proprietary formats to standard forms, and find mechanisms to render (e.g., format and then either display or print) files stored using those standards.

Regarding software, there is need to identify software that can be used at NDLTD sites, and to assist with the application of it to support NDLTD objectives. Regarding statistics and reporting, there is need to develop versions of surveys and other data-collection instruments that will work in various nations, and will allow fusion of findings across sites. Regarding

publisher relations, there is ample opportunity for a distributed approach to explain NDLTD to publishers and to enlist their support. In Germany this has proceeded well, with five professional societies (and publishers) serving as partners in their *Dissertations Online* effort. If each university were to involve all faculty who are editors of journals to obtain support from publishers for NDLTD, it would be relatively easy to change the current situation, in which many students are afraid to allow worldwide access because of concern over publisher reactions.

Other opportunities for joint work involve collaboration on developing supporting technology and on undertaking applied research to support the initiative.

4. Supported by Technology

The Virginia Tech team has worked on a variety of technologies that can support NDLTD. We explore some of these in the following subsections. Other technologies developed are discussed in [4].

4.1 PetaPlex™

Virginia Tech has purchased VT-PetaPlex-1, a system produced by Knowledge Systems Incorporated. This “superstorage” unit has 2.5 terabytes storage capacity. With 100 processors (each a Pentium II running at 233 MHz) running Linux, it has roughly 20 gigaflops computing capability as well. Virginia Tech offers to new members of NDLTD access to the PetaPlex system for archiving their ETDs.

4.2 MARIAN

Since 1990, Virginia Tech has been developing library search software called MARIAN that has evolved into a digital library system. Since 1993 [10] it has served as a research vehicle and alternative to the regular campus online public access catalog (OPAC), supporting searches against over a million MARC records. Numerous studies enhanced MARIAN with a graphical front-end as part of the ENVISION project [11-21]. With support from the National Library of Medicine, MARIAN is being converted to Java [22]. Its redesign has focused on scalability, flexibility, and reliability, so it may become an important tool to support NDLTD.

4.3 Federated Search

Virginia Tech hosts a federated search site that allows queries to be sent in parallel to some or all of the NDLTD sites that have searchable collections [23]. The underlying software handles resource discovery, multilingual query translation, and frame-based access to each system contacted. Sites with any of a number of search systems, or a Z39.50 interface, can be contacted. Extensions are planned to afford support for Harvest (in use by the German *Dissertation Online* project) and other search engines.

Significant improvements can be made to the federated search system. One key issue is how to integrate this with the Dienst software from Cornell and the NCSTRL project that makes use of Dienst [24].

4.4 Workflow Automation

The Library team at Virginia Tech that has been involved in NDLTD has developed and supports software to allow students to upload their works. This workflow management system also handles access by the Graduate School and the Library, including support for cataloging. It runs under UNIX, especially on a Sun, and uses MySQL database software.

Many other sites have uploaded this software and adapted it for local use. Some detailed effort is needed to make it work on other platforms and with other database management systems. That can be carried out in distributed fashion.

Planned enhancements to be undertaken at Virginia Tech include:

- standardizing on the metadata to be collected, following the Dublin Core, according to a draft specification developed mainly by Paul Mather in fall 1998;
- making use of SGML or XML submissions to extract needed metadata, eliminating the requirement for authors to provide that through forms;
- converting data collected into a form suitable for use with OCLC's SiteSearch;
- converting data collected into a form suitable for use with IBM Digital Library;
- providing mapping software to produce MARC records from the metadata; and
- integrating the annotation system developed by Todd Miller.

5. Research Challenges

NDLTD, like other digital library efforts, can benefit from a variety of research studies. Some involve moving forward the technologies explained in Section 4. Others fall into three main categories, covered in the next subsections.

5.1 Supporting Societies

Fundamentally, NDLTD serves various Societies, whose needs are highlighted in the discussion in Section 2. The various societies require support in developing ETDs, in submitting them, and in accessing the federated collections. Our software includes a database management system and scripts to support submission and workflow management. These also support browsing and support of various user interface routines. On the access side are commercial search engines, such as those from OpenText, OCLC, and IBM as well as our federated search software.

A broad range of additional software has been developed, mostly in the form of specialized tools and scripts. New members of NDLTD download this set as part of the process of developing a local support infrastructure.

Extensive additional work is required to construct the most effective interfaces, and to tailor interfaces to the various tasks carried out by various user groups. A variety of interfaces have been developed and tested [4], but additional work is required. The most effective long-term solution appears to further extend MARIAN to support all desired functions.

5.2 Preservation

A key concern for many who hear about NDLTD and Virginia Tech's decision to only accept electronic submission is that of long term preservation. While Virginia Tech Libraries accept responsibility for providing access and preservation services to the local collection, a fair amount of research is involved in digital preservation. Our aim is to explore this topic as needed for NDLTD to become successful. That means having a clear approach, validating it

at Virginia Tech and other interested places, and developing an economic model with accurate predictive capabilities. It relates to the selection of standards and related software, since suitable choices in those arenas can radically change the costs involved.

It will become much easier to convince others to move toward requiring (only) electronic submission once there are solid results regarding the challenging problem of preservation.

5.3 Education

Viewing NDLTD as an educational effort, one should focus on demonstrating its benefits regarding learning. Our research involves various instruments and observations to demonstrate project success. There are obvious counts like how many universities have joined NDLTD, how many ETDs are submitted each year, and how many universities are in each of the phases related to implementation of NDLTD at a local site. In addition, we evaluate each workshop, collect data from each student in conjunction with the submission process, and gather information from all cooperating users of Virginia Tech search software at the end of each session.

We hope to collect and relate findings from similar data from other members of NDLTD. Even harder may be to determine how NDLTD is used for each Society involved. We need to determine if the collection leads to classroom use by instructors and/or student access. We need to ascertain the effect of ETDs on other theses and dissertations. We need to collect data on how often ETDs are cited, relative to other genre. Ultimately we seek to determine who learns as a result of NDLTD, how, and what can be done to enhance learning.

6. Conclusion

NDLTD is a comprehensive digital library project involving over sixty members as of June 1999. We summarize three key issues in the following concluding subsections.

6.1 Need for Members

NDLTD had grown significantly, but still only involves a small percentage of the graduate degree granting institutions around the world. Others should join, to help their students, to ensure that the next generation of scholars is prepared for the Information Age, and to facilitate university collaboration.

6.2 Achieving Critical Mass

NDLTD is likely to grow rapidly once critical mass is achieved. That means having large universities that award many graduate degrees. It calls for leading institutions in each nation to participate, as has occurred in national projects in Australia and Germany. At that point large numbers of interesting works will be available, on every topical area. As the number of accesses increases, there will be added incentive to join.

6.3 Growth and Expansion

Though NDLTD has grown rapidly, it still has considerable room for improvement. Better tools, better training materials, more flexible software, more valuable services - all will promote such growth. Standards and common Web pages will promote interoperability and lower the cost per site, as well as facilitate long term involvement.

As NDLTD grows, some changes will be needed. The metadata standards, training resources, and federated search systems all need improvement to support more members and more users. The emerging committee structure will extend the reach of the Steering Committee to manage such growth. The annual workshop will serve as a vehicle for promoting sharing. Other meetings in connection with international digital library efforts will support this at the global level. Ultimately we hope that NDLTD will broadly support graduate education and research, extend collaboration among universities, and prepare the next generation of scholars for the Information Age.

7. Acknowledgements

The U.S. Department of Education, through its Fund for the Improvement of Post Secondary Education, P116B61190, for 1996-1999, has sponsored Virginia Tech co-PIs Edward A. Fox (project director), John Eaton, and Gail McMillan in their work on "Improving Graduate Education with a National Digital Library of Theses and Dissertations". Adobe, IBM, Microsoft, and OCLC have donated substantial hardware and software to help with NDLTD. Robert Akscyn and his company Knowledge Systems Incorporated have provided VT-PetaPlex-1 and collaborated on related research. Special thanks go to graduate students Neill Kipp, Paul Mather, and Constantinos Phanouriou. Many other faculty and students at Virginia Tech, and NDLTD members around the globe, also have contributed. Their assistance is gratefully acknowledged.

References

- [1] E. A. Fox, J. Eaton, G. McMillan, N. Kipp, L. Weiss, E. Arce, and S. Guyer, "National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources," *D-Lib Magazine*, vol. 2, 1996. <http://www.dlib.org/dlib/september96/theses/09fox.html>
- [2] E. A. Fox, J. L. Eaton, G. McMillan, N. Kipp, P. Mather, T. McGonigle, W. Schweiker, and B. DeVane, "Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources," *D-Lib Magazine*, vol. 3, 1997. <http://www.dlib.org/dlib/september97/theses/09fox.html>
- [3] E. A. Fox, R. Hall, N. A. Kipp, J. L. Eaton, G. McMillan, and P. Mather, "NDLTD: Encouraging International Collaboration in the Academy," *Special Issue on Digital Libraries of DESIDOC Bulletin of Information Technology (DBIT)*, vol. 17, pp. 45-56, 1997.
- [4] C. Phanouriou, N. Kipp, O. Sornil, P. Mather, and E. A. Fox, "A Digital Library for Authors: Recent Progress of the Networked Digital Library of Theses and Dissertations," presented at The Fourth ACM Conference on Digital Libraries, DL '99, Berkeley, CA, 1999.
- [5] M. Lesk, *Practical Digital Libraries: Books, Bytes and Bucks*. San Francisco: Morgan Kaufmann Publishers, 1997.
- [6] E. A. Fox and G. Marchionini, "Toward a Worldwide Digital Library; Guest Editors' Introduction to Special Section on Digital Libraries: Global Scope, Unlimited Access," *Comm. ACM*, vol. 41, pp. 28-32, 1998. <http://purl.lib.vt.edu/dlib/pubs/CACM199804>
- [7] ARL, *Electronic Theses and Dissertations*, vol. 7, Spec Kit 236 ed. Washington, D.C.: Association of Research Libraries, 1998. <http://www.arl.org/transform/>
- [8] E. A. Fox, G. McMillan, and J. Eaton, "The Evolving Genre of Electronic Theses and Dissertations," presented at Digital Documents Track of HICSS-32, Thirty-second

- Annual Hawaii International Conference on Systems Sciences (HICSS), Maui, HI, 1999. <http://scholar.lib.vt.edu/theses/presentations/Hawaii/ETDgenreALL.pdf>
- [9] E. A. Fox, N. Kipp, and P. Mather, "How Digital Libraries Will Save Civilization," *Database Programming & Design*, vol. 11, pp. 60-65, 1998. <http://www.dbpd.com/foxweb.html>
 - [10] E. Fox, R. France, E. Sahle, A. Daoud, and B. Cline, "Development of a Modern OPAC: From REVTOLC to MARIAN," in *Proc. 16th Annual Int'l ACM SIGIR Conf. on R&D in Information Retrieval, SIGIR '93*. Pittsburgh: ACM Press, 1993, pp. 248-259.
 - [11] E. A. Fox, D. Hix, L. Nowell, D. Brueni, W. Wake, L. Heath, and D. Rao, "Users, User Interfaces, and Objects: Envision, a Digital Library," *J. American Society Information Science*, vol. 44, pp. 480-491, 1993.
 - [12] E. A. Fox, N. D. Barnette, C. Shaffer, L. Heath, W. Wake, L. Nowell, J. Lee, D. Hix, and H. R. Hartson, "Progress in Interactive Learning with a Digital Library in Computer Science," in *ED-MEDIA 95, World Conference on Educational Multimedia and Hypermedia*. Graz, Austria, 1995, pp. 7-12.
 - [13] L. Heath, D. Hix, L. Nowell, W. Wake, G. Averboch, and E. A. Fox, "Envision: A User-Centered Database from the Computer Science Literature," *Communications of the ACM*, vol. 38, pp. 52-53, 1995.
 - [14] L. Nowell and D. Hix, "User interface design for the project Envision database of computer science literature," in *Twenty-second Annual Virginia Computer Users Conference*. Blacksburg, VA, 1992, pp. 29-33.
 - [15] L. Nowell and D. Hix, "Visualizing search results: User interface development for the project Envision database of computer science literature," in *Advances in Human Factors/Ergonomics, Proceedings of HCI International '93, 5th International Conference on Human Computer Interaction*, vol. 19B, *Human-Computer Interaction: Software and Hardware Interfaces*. Elsevier, 1993, pp. 56-61.
 - [16] L. Nowell and D. Hix, "Query composition: Why does it have to be so hard?," in *East-West International Conference on Human-Computer Interaction*, vol. I. Moscow, Russia, 1993, pp. 226-241.
 - [17] L. Nowell, E. A. Fox, L. Heath, D. Hix, W. Wake, and E. Labow, "Seeing Things Your Way: Information Visualization for a User-Centered Database of Computer Science Literature," Virginia Tech Dept. of Computer Science, Blacksburg, VA Technical Report TR-94-06, January, 1994.
 - [18] L. T. Nowell and E. A. Fox, "Envision: Information Visualization in a Digital Library". Demonstration. Seattle, WA: ACM SIGIR'95, July 10, 1995.
 - [19] L. Nowell, D. Hix, R. France, L. Heath, and E. A. Fox, "Visualizing Search Results: Some Alternatives to Query-Document Similarity," in *SIGIR '96*. Zurich, Switzerland, 1996, pp. 67-75.
 - [20] L. T. Nowell, R. K. France, and E. A. Fox, "Visualizing search results with Envision". Demonstration. Zurich, Switzerland: ACM SIGIR'96, Aug. 19, 1996.
 - [21] L. Nowell, "Graphical Encoding for Information Visualization: Using Icon Color, Shape and Size to Convey Nominal and Quantitative Data," Virginia Tech Dept. of Computer Science, Blacksburg, VA, Ph.D. Dissertation, 1997.
 - [22] J. Zhao, "Making Digital Libraries Flexible, Scalable, and Reliable: Reengineering the MARIAN System in JAVA," Virginia Tech Department of Computer Science, Blacksburg, VA, Master of Science, 1999.
 - [23] J. Powell and E. Fox, "Multilingual Federated Searching Across Heterogeneous Collections," *D-Lib Magazine*, vol. 4, 1998. <http://www.dlib.org/dlib/september98/powell/09powell.html>

- [24] C. W. Sharrets and J. C. French, "Electronic Theses and Dissertations at the University of Virginia Library," presented at The Fourth ACM Conference on Digital Libraries, DL '99, Berkeley, CA, 1999.