# ETDs, NDLTD, and open access: a 5S perspective

**Edward A. Fox**

Executive Director, NDLTD; professor.

**Seungwon Yang**
PhD Student.

**Seonho Kim**
PhD Student.

Department of Computer Science, Virginia Tech, Blacksburg, VA 24061 USA.
http://www.cs.vt.edu
E-mail: fox,seungwon,shk@vt.edu

## Abstract

*Worldwide initiatives toward digital library (DL) support for electronic theses and dissertations (ETDs), facilitated by the work of the Networked Digital Library of Theses and Dissertations (NDLTD), are a key part of the move toward open access. When all graduate students learn to use openly available ETDs, and have experience with authoring and submission in connection with their own research results, it will be easy for them to continue these efforts through other contributions to open access. When all universities support ETD activities, they will be key participants in institutional repositories and open access, and will have engaged in discussion and infrastructure development supportive of further open access activities. Understanding of open access also can be facilitated through modeling of all of these efforts using the 5S framework, considering the key aspects of DL development: Societies, Scenarios, Spaces, Structures, and Streams.*

### Keywords

*5S (societies, scenarios, spaces, structures, streams). Curricula. DL (digital library). ETD (electronic thesis or dissertations). NDLTD (networked digital library of theses and dissertations). Open access. OAI (open archives initiative). Standards.*

## ETDs, NDLTD e acesso aberto: uma perspectiva 5S

### Resumo

*Iniciativas internacionais para o suporte de teses e dissertações eletrônicas (ETDs) através de bibliotecas digitais (DL), facilitadas pelo trabalho da Biblioteca Digital em Rede de Teses e Dissertações (NDLTD), são um fato chave no caminho ao acesso aberto. Quando os alunos de pós-graduação aprenderem a usar as ETDs disponíveis e tiverem experimentado a criação e a submissão dos trabalhos resultantes de suas pesquisas, ele serão participantes ativos nos repositórios institucionais e no acesso aberto. Ao mesmo tempo, poderão se engajar nas discussões e na criação de infraestrutura que suporte o crescimento do acesso aberto. A compreensão do acesso aberto pode ser facilitada pela modelagem 5S aplicada aos aspectos fundamentais das biblitoecas digitais:* Societies *(Sociedades),* Scenarios *(Cenários),* Spaces *(Espaços),* Structures *(Estruturas) e* Streams *(Correntes).*

### Palavras-chave

*5S (societies = sociedades, scenarios = cenários, spaces = espaços, structures = structures, streams = correntes). Currículos. DL (digital library = bibliotecas digitais). ETD (electronic thesis or dissertations = teses ou dissertações eletrônicas). NDLTD (networked digital library of theses and dissertations = biblioteca digital em rede de teses e dissertações). Acesso aberto. OAI (open archives initiative = iniciativa dos arquivos abertos). Padrões.*

## INTRODUCTION

One of the easiest and most effective ways to promote open access to research and educational content involves support of electronic theses and dissertations (ETDs) – covering their: authoring, submission, workflow processing, storage, archiving, harvesting, discovery, reading, and referencing. As a result of such support, in 2006 there is widespread discovery and use of the hundreds of thousands of ETDs already freely accessible, from every continent around the globe, covering all topical areas. Advocacy activities, community building, development of standards, documentation of best practices, and other assistance for ETD initiatives are coordinated by the Networked Digital Library of Theses and Dissertations (NDLTD) [25]. It relies upon worldwide engagement of university students, faculty, and staff (especially those involved in libraries or graduate programs) – as well as support from corporations and initiatives operating at regional and national levels, typically related to open access, networking, library/information science, and/or graduate education.

The first author of this paper became interested in this in 1987 [34], when UMI hosted a workshop in Ann Arbor, Michigan, USA to explore how the Electronic Manuscript Project [69], which was based on using SGML [42] for electronic publishing, might relate to doctoral dissertations. It is our hope that by the 20th anniversary of that discussion, the number of ETDs archived in a single year will exceed 100,000. By then, we fully expect ETDs to be one of the most important genre in the unfolding of the open access movement [40]. This article summarizes progress and plans in that direction.

## DIGITAL LIBRARIES

NDLTD is part of the movement toward digital libraries (DLs). Early visions of digital libraries (related to, and sometimes called: content management systems, digital repositories, electronic libraries, institutional repositories, knowledge management systems, or virtual libraries) date back to the 1960s [60] and before. Enabling research to promote development of DLs received initial funding in USA [31] and other nations in the 1990s, in part as a result of efforts of those in the computing,
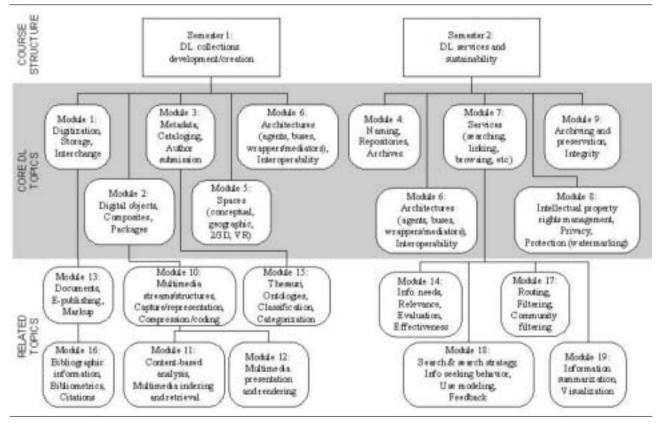
library, and information science fields to explore the synergies and applications of decades of fundamental investigations [29]. By 1993 the DL field was perceived as a hot topic [28]. By 1996 there were major annual professional conferences in the field [39] leading to even larger coordinated events for the Americas [11], Europe [22], and Asia [51], along with workshops and national/regional events.

There are many publications in the DL area, including brief overviews [27] and longer reviews [41]. There are online magazines [18] and journals [52]. In connection with a DL curriculum development project funded for 2006-2008, we examined a substantive sample of the DL magazine and conference paper literature to help us identify what topics relate to DL, and which of those topics might be considered "core". Figure 1 summarizes our first attempt to identify the key topics that relate to DL, to specify what topics might be covered in a DL "knowledge module" (typically corresponding to a portion of a course), and to suggest how these might fit into various DL curricula. We identified 9 modules we considered core, numbered 1-9, shown in the middle of the figure. Since these were most important for our work

on DL curricula and educational resources, we refined these into the set shown in Table 1. Then we were able to study the DL literature and manually classify works according to those revised 9 topics (modules). Figure 2 shows the topical coverage of topics 1-9 for *D-Lib Magazine* [18]. The colors reflect year of publication, so for each topic it is possible to perceive the evolution, and to note shifts in degrees of coverage over the years. By way of comparison, Figure 3 shows the topical coverage for papers in DL conferences. Figure 4 is similar, but instead of classifying papers, we classified sessions (i.e., small groups of papers presented together) at DL conferences.

These figures suggest that there has been significant accomplishment by those in the DL community, and relatively rapid movement in the directions recommended in the early 1990s. Further progress will be assured if research and development activities are supported by adequate funding programs, and guided by standards and other types of community agreement. Since two key goals of NDLTD are to help advance the digital library field, and to ensure that graduate students become knowledgeable about DLs, we aim to help
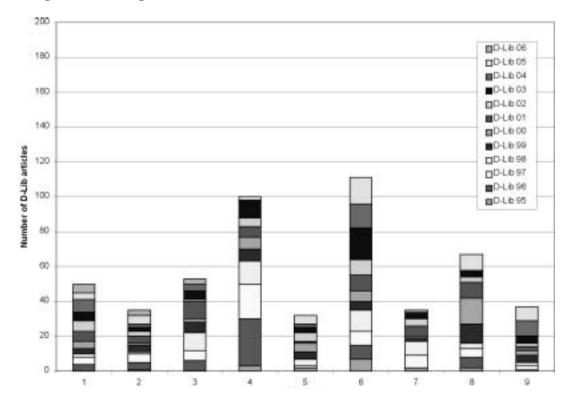
FIGURE 1
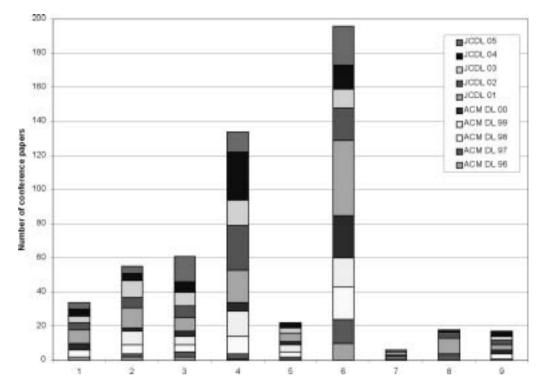**Initial set of curricular modules for DL topics**

TABLE 1
**Digital Curriculum Module Scopes**

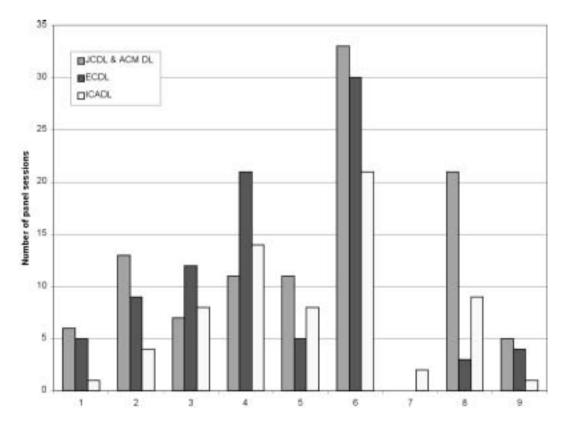| N° | Title | Content Details |
|----|-------|-----------------|
| 1 | Collection Development | Digitization<br>Document and E-publishing Markup<br>Harvesting |
| 2 | Digital objects / Composites / Packages | Text Resources<br>Multimedia streams/structures, Capture/representation, Compression/coding: content-based analysis, multimedia indexing, multimedia presentation rendering |
| 3 | Metadata, Cataloging, Author submission | Thesauri, Ontologies, Classification, Categorization<br>Bibliographic information, Bibliometrics, Citations |
| 4 | Architecture, Interoperability | Agents, buses, wrappers/mediators |
| 5 | Spaces (conceptual, geographic, 2/3D, VR) | Storage<br>Repositories, Archives |
| 6 | Services (searching, linking, browsing, etc.) | Info needs, Relevance, Evaluation, Effectiveness<br>Search & search strategy, Info seeking behavior, User modeling, Feedback<br>Routing, Filtering, Community filtering<br>Sharing, Networking, Interchange<br>Info summarization, Visualization |
| 7 | Intellectual property rights management, Privacy, Protection (watermarking) | Defines the purpose of copyright and copyright protection of DL resources<br>Discusses the controversial issues related to privacy<br>Deals with technical methods to protect the authorship of resource creators |
| 8 | Social issues / Future DLs | Related to DL design and development for a specific group of users or of particular topics, and future DL descriptions or projections |
| 9 | Archiving and preservation integrity | Long-term plans for digital resource preservation, migration, emulation, etc.<br>Fundamental strategies to preserve digital resources, preservation models |

FIGURE 2
**Topical coverage in the DL magazine literature**

FIGURE 3
**Topical coverage in the DL conference literature**



FIGURE 4
**Topical coverage for sessions in DL conferences**

especially in these regards. The sections below highlight progress and plans – showing how they contribute to open access, and describing them by way of a framework that also might be of use when describing other DL-related activities.

## STANDARDS

**Internet, WWW:** One of the key foundations of success in the information, computing, and communications area has been the development of appropriate standards. The Internet was based on communication standards like TCP and IP, and a growing number of protocols for services like FTP and SMTP. The WWW was built upon information standards like HTML and XML, naming agreements such as URLs and DOIs, and protocols supported by web servers and browsers. Here we summarize key standards related to DLs (especially those applicable to the efforts of NDLTD), many of which have facilitated the movement toward open access.

**Content and Formats:** Since the DL field is so broad, in describing DL standards we elect for the sake of brevity to focus on those that relate strongly to ETDs. If we start with the actual content, the most popular are PDF and XML. Though the earliest interest in ETDs arose from consideration of SGML, and though some ETDs have been prepared (dating back to 1988) in accordance with that standard, the cost of suitable authoring tools and training made widespread use of SGML for ETDs problematic. Fortunately, XML has many of the same advantages of SGML, supporting descriptive markup and even more flexible rendering (e.g., with XSLT), so has been the cornerstone of a number of ETD initiatives (e.g., in Chile and France).

But ETDs are prepared by students, often with somewhat narrow experience in the areas of word processing and electronic publishing, and most students follow community practice in using Microsoft Word or similar programs for authoring. However, since ETDs represent archival publications, and in many regions must be kept accessible for 50 years or more, Word is generally not acceptable as the sole representation of a work. Fortunately, in the early 1990s, as the DL field was unfolding, and the WWW was emerging, PDF appeared as a format that rapidly became popular for preserving the rendered form and appearance of electronic documents [59, 8]. While from the earliest days of the move toward NDLTD it was agreed that having both a logical/descriptive (e.g., SGML or XML) and a rendered version (e.g., PDF) of ETDs was desirable, in most

universities the expedient choice was made to launch efforts with a focus on use of PDF, along with on-demand support for those interested in using XML. Fortunately, NDLTD has goals of supporting continuously improving education and training for graduate students, and empowerment of universities to move forward in adoption of the most effective technologies, so we see this matter as an area for continuous improvement rather than a source of contention.

**Multimedia:** Another area of improvement regarding ETDs is the extension of content types to go beyond simple text. Electronic publishing facilitates inclusion of color diagrams and figures, photos (typically using the JPEG standard [53] now commonplace in digital cameras), animations, audio files (now even more accessible through streaming servers and podcasts, usually as MP3 but sometimes using other formats), and videos (often as MPEG-1 or MPEG-2 [15]), depending on the quality of capture and the need to communicate precisely). In addition, there are many special formats employed for a wide variety of sensors and instruments, including those related to medical and health care, as well as science and engineering. Long lists of recommended standards have been prepared over the years, but practices shift widely. For example, some students prefer to work with state-of-the-art technology while others wish technology would disappear and let them focus on their research. Accordingly, policies lack consistency. For ETDs, where it is important to consider preservation for the long term, it may be best to be conservative and work with well established international standards (at least for the core of a thesis). The result should be that large numbers of future readers will benefit from the thoughtfulness of authors. Future readers also can benefit from multimedia inclusions in ETDs, aimed at communicating scholarly discoveries in the most effective manner, which can result from the author's willingness to supplement their subject matter expertise with skills regarding multimedia presentation and archiving. When new technologies are employed, as a hedge against the future, such supplemental files can be made available in a number of formats, so that at least one version is likely to be supported years later.

**Datasets:** Another emerging extension of ETDs deals with datasets. This is a hot topic in the e-science world, but clearly has wider scope. Many researchers are becoming aware that future advancement of knowledge, not just in science but more broadly across all areas of scholarship, depends upon long-term preservation and access support for raw data. Since more and more of that

data is available only electronically, and since some if not all versions of remaining data collections also have digital representations, keeping electronic datasets for the long term is crucial if theses and dissertations are to lead to validation or follow-on research. Some of these datasets are managed by government or professional libraries and archives, in which case students may deposit their data and simply keep a pointer or identifier in their ETD. But all too often, the preservation of datasets is left to the good auspices of students, faculty, or research groups/centers, which typically have little expertise or financial support for this task. Fortunately, university support for ETDs can easily be extended to facilitate dataset preservation, if the datasets are stored together with a submitted ETD, or are uploaded at roughly the same time to some separate local repository. It is strongly encouraged that universities institute policy recommendations in this regard (in keeping with disciplinary practices and with legal and economic procedures and decision making processes related to management of intellectual property), typically in conjunction with records management or library and archiving activities, preferably as part of the organization's information infrastructure. Here again, work with ETDs can be a driver for local efforts to enhance university support for research, and to increase involvement in discussion about long term needs.

**Naming, hypermedia, and superimposed information:** Naming is an important role for the discoverer. Having persistent names, which can be effectively used for the long term to connect with named entities, is another key part of the emerging global information infrastructure (cyberinfrastructure). Scholars have long faced these problems, now made visible as a result of the widespread use of (ever-changeable) URLs instead of URNs, URIs, DOIs, or other types of stable resource identifiers in the WWW [78]. Persistent names are needed for ETDs, for related datasets, for multimedia files connected with ETDs, and for other electronic content described in ETDs. In addition to having means to refer to such digital objects, it is desirable to refer to parts of such objects (e.g., a word, phrase, sentence, excerpt, paragraph, page, section, table, or figure in a publication; a face in a group photo; a tumor outline visible in an X-ray image; a theme being studied in a musical composition; a step in a procedure documented in a video). Hypermedia systems may provide such assistance, but often that is hard to sustain into the future. In conjunction with XML documents, there are schemes like XPath that provide appropriate functionality. More generally, markup schemes, like

those developed for various classes of documents through the work of the Text Encoding Initiative, provide genre-specific aids. Gradually, as efforts mature, for example the superimposed information middleware work based at Portland State University [73], it will be straightforward to "mark" (parts of) objects in an effective and persistent manner.

**Metadata:** Another type of supplement to an ETD is a metadata record (which may include some or all of the typical types of metadata such as descriptive, administrative, and structural). Older scholars will recall card catalogs, wherein a card (or several if different types of organizational schemes were employed, based on title, author, and category) described each work in the library collection. As these moved to electronic versions, standards like MARC-21 [24] emerged as the main approach to connect author, date, publisher, categories, keywords, and other attributes to document the provenance and to facilitate discovery of the work. While full-text indexing supports a (perhaps better) way to search for an ETD, searching with full-text plus metadata (plus citation and other supplemental information, possibly including content-based retrieval tailored to audio, image, and/or video content) is even better. Accordingly, the Dublin Core Metadata Initiative [17] emerged to develop metadata standards for electronic resources [100]. Thus, if an ETD has no MARC-21 or similar standard metadata record that has resulted from local library processing, at the very least there should be a Dublin Core [101] record with as many as appropriate of the standard 15 elements (i.e., fields or attributes) specified [19]. Even better, there should be an extended Dublin Core record, which also has elements of special important for theses and dissertations. Toward that end, and as a result of over three years of international meetings and discussions, in 1997 ETD-ms, the first ETD metadata standard was developed under NDLTD auspices [5]. The NDLTD Standards Committee is revisiting this work to extend it based on a decade of experience with ETD collections and a broader international perspective on needs and terminology.

For NDLTD-supported resource discovery of ETDs from around the globe, such a standard is especially valuable. However, it is expected that university, national, and regional standards for metadata also will exist because of local needs. Crosswalks from those metadata standards to ETD-ms will allow local and global situations to evolve in parallel for the widest benefit.

Ultimately, however, the quality of metadata about ETDs will depend on the training of authors to understand that describing their works is a responsibility of document creators. But, at least for the foreseeable future, there also is need for assessing that quality, improving that training, and supplementing the work of authors with the efforts of catalogers (metadata librarians) and other professionals. University librarians have an important role to play in these activities.

**Harvesting:** In addition to content-related standards, standards for communication protocols also have been important for ETDs. Theses and dissertations are produced in a decentralized manner, by graduate students attending thousands of colleges and universities around the globe. Their local institution is obliged to keep copies, and in some cases policies preclude putting copies in the libraries or collections of other organizations, so some means of dissemination is needed that involves the home institution.

The direct dissemination of actual works is feasible from home institutions if each ETD has a persistent name (e.g., URI), and that name is known to an interested party. But discovery of relevant works, which each lead to a persistent name, typically requires some communication scheme.

One such popular scheme involves crawling. This is the method employed by search engines, such as Google. However, ETDs are not always found during a crawl, and search engines may have trouble in provided coordinated access to the various pieces and related files of an ETD (e.g., when each chapter and multimedia attachment is in a separate file). Crawling does not locate works in the Deep Web [65]. Those works are more amenable to finding through federated search or harvesting.

Federated search is supported by the international standard Z39.50 [64]. Universities or regional services that store metadata about ETDs can index their local content and handle queries sent using the protocol for Z39.50. All sites of interest can be searched in parallel, and the results can be merged for each query by a server or client program. However, as the number of sites being searched increases, performance often degrades relative to other approaches [3]. Furthermore, quality (e.g., with regard to performance, metadata completeness, presentation of results, and search functionality) is limited by the least helpful of all the sites in a federation that supports Z39.50.

Accordingly, many distributed services like NCSTRL have shifted toward "harvesting" as a more appropriate way to support communities of users [3]. The most popular de facto standard for harvest-based services was developed by 2000 [98] as part of the Open Archives Initiative (OAI) [97]. A site that maintains a catalog of ETDs can expose the metadata in that catalog by running software that supports the Protocol for Metadata Harvesting (OAI-PMH) [58]. Services like ARC [61], which tries to find all OAI "data providers" and harvest their metadata into a single collection that covers a wide variety of sources (e.g., individual to global collection, with location or topic based scope) and genres (e.g., e-prints, pre-prints, bibliographies, student works, educational resources, reports), have very broad coverage.

There are many ways that OAI can be used to support work with ETDs [92]. First, if a university wishes to share its metadata with regard to its collection of ETDs, it can select any of a number of software systems to help. The simplest and most focused is etd-db [56], which grew out of efforts at Virginia Tech in the late 1990s, and is being enhanced further in 2006. But OAI repositories can have a "set" structure imposed atop the collection of metadata records, so institutional repositories (e.g., DSpace [71]), that aim to collect all of the types of documents prepared at a university, can have a separate set for the local ETDs. Then a harvester interested only in ETDs, when connecting with an institutional repository, can restrict its request (for new works) to those in the ETD set.

Second, since NDLTD encourages universities to enhance their DL-related infrastructure, it is appropriate that they learn to test their ETD data provider with the OAI Repository Explorer [88]. This can help ensure that others can harvest desired data.

Third, and the most visible way that OAI connects with ETDs, it is helpful to develop union catalogs. These can be built using suitable harvesting procedures. One was launched in 2001 by NDLTD. In 2003, management of catalog was shifted from Virginia Tech to OCLC (acting on behalf of NDLTD) [93]. The NDLTD Union Catalog run by OCLC [75] included 257K records from 60 data providers as of July 2006. It is hoped that, as use of the union catalog grows, and more and more services are built atop it, more universities will support the OAI, thereby greatly facilitating the discovery of their ETDs.

**Logging:** Another area where standardization can be of benefit for DLs is with regard to data collection, analysis,

and evaluation. It is difficult to assess to what degree collections of ETDs are popular, to find which ones are most desired, to contrast the use of metadata records vs. full-text vs. multimedia files, or to ascertain which parts of the internet have the largest numbers of readers interested in ETDs. One source of data in this regard is from DL logs; we have proposed a standard in that regard [47]. Hopefully the DL community, or subsets of it like those working with ETDs, will log similar data about DL system operation and user behavior, so local and aggregate statistics can be produced. These can provide helpful insights, as will be seen below with regard to the discussion of Figures 6 and 7.

## DOCUMENTATION

Thousands have been involved in the unfolding of the worldwide ETD initiative. Discussions have proceeded since 1987, with early events discussed in a series of articles in *D-Lib Magazine*. The 1996 paper covers early efforts, included US activities funded starting in 1995 by SURA and the US Department of Education [34]. The NDLTD acronym was retained in 1997, when the first word in the long name shifted from "National" to "Networked" [35], indicating a broadening of scope to serve the international community. A 1998 *D-Lib* paper showed how multilingual access was supported by a federated search system [79]. A two-part *D-Lib* series appeared in 2001 to summarize progress, including the shift from federated search to harvesting to support searching [91, 90]. Later that year a paper appeared about Open Digital Libraries (ODL, see [94]), a scheme to support a component-based approach to DL construction, which was deployed to facilitate searching of the NDLTD Union Catalog.

In 2004 Marcel Dekker published a book about ETDs [36], to supplement its other works to support scholars. This edited volume covers a broad range of international perspectives regarding ETD initiatives. It considers the concerns of students, faculty, libraries, graduate schools, administrators, and technologists. There is discussion of intellectual property and copyright, of PDF and SGML/XML, and of novel modes of expression that involve multimedia and hypermedia. Such innovation by ETD authors has been encouraged in recent years by an award program sponsored by Adobe. Adobe also has a website with documentation about ETD activities. Adobe funded the development of tutorial materials to help authors who are creating ETDs in PDF [1].
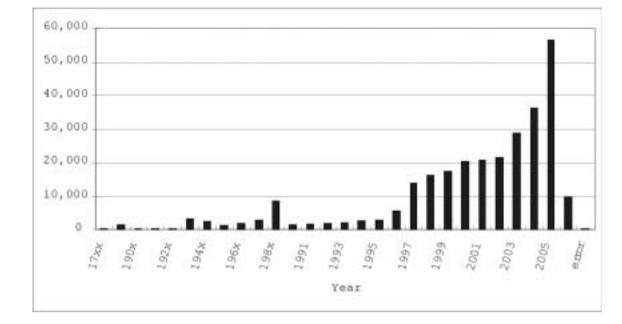
There also is an online book about ETDs, originally funded by UNESCO, in multiple languages (e.g., English, French, Greek, Spanish). The *ETD Guide* [72] was the result of an international collaboration, with contributors from, e.g., Australia, Brazil, Canada, Chile, France, Germany, and USA. Work on the Guide was launched in part as a result of a 1999 workshop at UNESCO headquarters in Paris [77]. NDLTD plans to provide updates to the Guide, initially coordinated through a wiki.

Further documentation about ETD initiatives has appeared through the proceedings of a series of international conferences on this topic [74]. NDLTD has been the key sponsor. Recent meetings have been in Germany (2003), USA (2004 [57]), Australia (2005), and Canada (2006). Meetings in 2007 and 2008 will be in Sweden and the United Kingdom.

An easy way to obtain information about ETDs is from the NDLTD site [25]. In addition to documentation, information about membership and committees, and links to conference announcements and publications, one can select any of a number of services to facilitate searching and browsing. Virginia Tech supports one service based on ODL [94]; a mirror version adapted for the Chinese language is hosted by CALIS in Beijing [14]. Additional search services are run by VTLS [99] (with versions of the interface, and metadata records, in a number of languages) and Scirus [23] (with full-text indexing). Discussions are underway with a number of search engine sites (e.g., Google, Microsoft) to provide additional services to help ensure broader use of ETDs worldwide.

Virginia Tech also runs an experimental system, operating atop the search system by FAST. Seonho Kim has been logging and analyzing activity with that system [55]. For example, Figure 5 shows his reporting of the growth of the number of ETDs based on their date of creation. It is likely that numbers will continue to rise rapidly in upcoming years, as more and more institutions launch ETD initiatives, and as existing initiatives mature and lead to more aggressive policies on submission of ETDs to a local repository. When submission (which is different from providing access) is required, the numbers go up quite rapidly! Some institutions also have retrospective conversion programs to digitize older works, either when they are requested, or as a comprehensive effort (as is being done at Virginia Tech now); these also help increase the number of ETDs available. We look forward to when the NDLTD Union Catalog has more than a million records, and has hundreds of thousands of works added each year.

FIGURE 5
**Growth in numbers of ETDs**



## 5S CHARACTERIZATIONS

Since worldwide activities with regard to ETDs are diverse, since NDLTD's efforts to support these activities are varied, and since open access relates to a large number of issues, it is important to have a powerful framework in which to characterize the situation. Since 1999 we have been developing just such a formal framework for DLs [30]. Key aspects of our 5S framework are summarized in Table 2.

The 5S framework is particularly applicable to DL modeling. It has been used for a variety of case studies, such as to model DLs for archaeological sites as well as regional and global DLs built by harvesting from the local DLs [85]. Two case studies were undertaken in 1999 to explore the use of 5S for describing educational DLs [32]. These covered educational resources for computing, and ETDs. A 2004 case study focused on ETDs was based in Brasilia [80].

A good summary of 5S, including how it can be used to describe ETD activities, appeared in 2004 [45]. It draws in large part on the dissertation work of Gonçalves [44]. The 2006 dissertation by Shen [84] builds on this, adding in key results related to quality, interoperability, and integrated support for various types of exploration (e.g., searching, browsing, and visualization). Future work on global ETD services, considering the increasingly sophisticated regional and national efforts in the

Americas, Australasia, and Europe, could benefit from the advances in 5S made by Gonçalves and Shen.

Modeling the **Societies** that relate to a DL is of particular import, from a 5S perspective. In the case of ETDs, there clearly are many considerations in this regard. At the broadest, we have an international community that is moving toward tighter collaboration, across space (leading to a global consciousness) and time (involving old as well as new ETDs, and involving students new to the world of research, as well as those with extensive publication experience) [38]. A key Society is that of people involved in graduate education [21]. New to that scene are the authors of ETDs, who need various kinds of support [76]. But they also are the ultimate innovators, who will make sure that the genre of ETDs develops and matures, allowing them to communicate ever more effectively [40]. While some critics have suggested that students would feel burdened if required to work with ETDs, for most students this is a non-issue. A variety of surveys have shown that students generally are favorably disposed toward ETDs; in reality there are no serious problems [2] [20]. Indeed, if one considers that theses and dissertations are the main, and sometimes the only artifact resulting from years of student labor, and that having ETDs available may increase the number who read them by a factor of 100 or 1000, students are among those with the most to gain from ETD initiatives [66]. They also can gain when there is strong support for ETD authors [76].
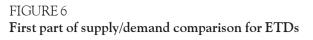
TABLE 2
S overview

| Ss | Examples | Objectives |
|---|---|---|
| **Streams** | Text; video; audio; image | Describes properties of the DL content such as encoding and language for textual material or particular forms of multimedia data |
| **Structures** | Collection; catalog; hypertext; document; metadata | Specifies organizational aspects of the DL content |
| **Spaces** | Measure; measurable, topological, vector, probabilistic | Defines logical and presentational views of several DL components |
| **Scenarios** | Searching, browsing, recommending | Details the behavior of DL services |
| **Societies** | Service managers, learners, teachers, etc. | Defines managers, responsible for running DL services; actors, that use those services; and relationships among them |

Modeling the **Societies** that relate to a DL is of particular import, from a 5S perspective. In the case of ETDs, there clearly are many considerations in this regard. At the broadest, we have an international community that is moving toward tighter collaboration, across space (leading to a global consciousness) and time (involving old as well as new ETDs, and involving students new to the world of research, as well as those with extensive publication experience) [38]. A key Society is that of people involved in graduate education [21]. New to that scene are the authors of ETDs, who need various kinds of support [76]. But they also are the ultimate innovators, who will make sure that the genre of ETDs develops and matures, allowing them to communicate ever more effectively [40]. While some critics have suggested that students would feel burdened if required to work with ETDs, for most students this is a non-issue. A variety of surveys have shown that students generally are favorably disposed toward ETDs; in reality there are no serious problems [2] [20]. Indeed, if one considers that theses and dissertations are the main, and sometimes the only artifact resulting from years of student labor, and that having ETDs available may increase the number who read them by a factor of 100 or 1000, students are among those with the most to gain from ETD initiatives [66]. They also can gain when there is strong support for ETD authors [76].

Such support, however, typically only occurs when there is active leadership and support for change in the local university community [37]. The amount and level of such leadership is a key determiner of how quickly an effective ETD program can be put in place. In many universities, launching an ETD program, and evolving it so that students are required to submit works, may take several years. But with strong high level support, the whole process can be completed in half a year [49]. This is getting easier as time goes by, since effective practices and policies are well known and have been reported [26]. There also is a growing cadre of people with experience in implementing successful ETD programs, a strong commitment to mentoring, and collaboration between more and less developed nations [87].

One other fortunate situation with regard to ETD programs, relating authors and readers, is a rough balance in supply and demand. Seonho Kim studied this, using works in ETD collections to characterize supply, and query logs to characterize demand [55]. To provide a context for comparison, he used 77 different topical categories, and classified ETDs and queries based on those categories. Figures 6 and 7 show the results for those 77 categories. Though for many topics an approximate balance exists, for a small number of categories – perhaps good topics for future research – there is more demand than supply.
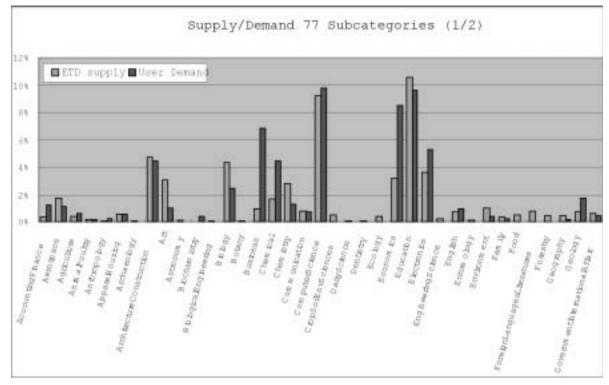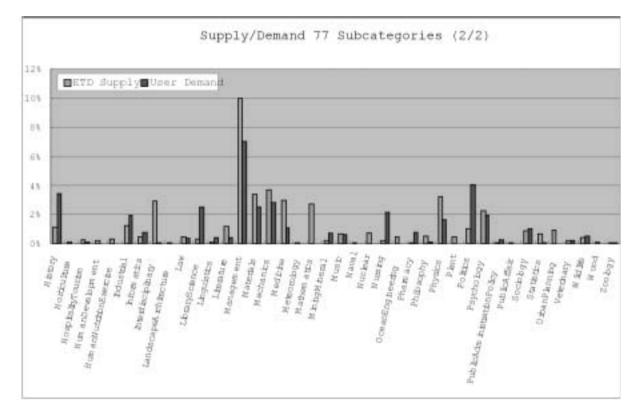
FIGURE 6
**First part of supply/demand comparison for ETDs**



FIGURE 7
**Second part of supply/demand comparison for ETDs**

Modeling the **Scenarios** that relate to NDLTD leads to a discussion of services provided, through systems, by institutions. Fundamental are those that help with local activities [68, 67]. Typically, libraries, else computing / information technology centers, manage those services. Clearly they are the most appropriate to devise and enforce policies, support authors, certify quality, operate institutional repositories, and facilitate long term archiving and preservation. However, other parties must be involved if those seeking helpful research works in a global context are to find the right ETDs from among the collections of many thousands of educational institutions.

One type of institution with interest in supporting access is the national library. Borbinha, discussing activities at the Portuguese National Library, argued in 1998 for federated access and services [10]. To help suit the needs around the globe, multilingual federated search was tested at Virginia Tech, starting in 1998 [79]. Besides functionality, however, usability also is a key consideration regarding services. A 1999 usability study of several digital libraries, both commercial and open source, covering both proprietary and open access collections, found the NDLTD services acceptable, but also highlighted areas for improvement (for all systems tested) [54]. Consequently, a range of services have been developed, as discussed near the end of the prior section.

Many additional services could be offered. A 1999 study through focus groups, with an accompanying pilot study, made clear that annotation services are of interest [54]. Improved methods for resource discovery, search, browsing, etc. could be of help [48] [63]. There is almost a complete void with regard to potential support for multimedia content-based access [70]. Richardson has been working on a promising approach to multilingual summarization and resource discovery by way of concept maps accompanied by machine translation (that makes use of identification of parallel corpora) [82, 83]. As these and other services develop, they can be added to component pools [89]. Components can be brought together in DLs, or, if made available through a service oriented architecture, can help in the move toward the Semantic Web [7].

Services also help with the integration of ETDs in the Web infrastructure. Ultimately we hope that all ETDs will be harvested using OAI-PMH, so there can be a comprehensive Union Catalog [95]. However, some institutions lack expertise with that protocol, and are used to just putting up works on the WWW, with the expectation that crawlers will find them and help provide access. Though they may be right, not all services will pick up ETDs in their entirety, and fewer still will support search that utilizes both metadata and full-text indexing. One promising scenario to deal with these challenges is to construct a DL by semi-automatically identifying small ETD collections on the Web [13]. We have demonstrated that the Web-DL approach [13] can help in this regard, but a fair amount of work is involved, which may not be feasible for a light-weight organization like NDLTD.

Scenarios by default are based on an assumption of quality. In real life, however, high quality services are difficult to build and maintain, so focusing on quality is not universal. But DL quality [96] is a key issue for NDLTD [33], since we hope to attract new authors and readers, and to ensure they are comfortable life-long users. Thus, NDLTD is one of the few DL organizations that considers the entire information life cycle. Hence, it was possible to assess a number of indicators of DL quality by studying the content connected with the ETD Union Catalog [44]. Working with a range of indicators, one can fit them into models to help predict intention to (re)use a DL [84]. We hope ultimately to have a more comprehensive view of DL quality, and to facilitate support of broad communities of those working with DLs [46]. These then can be extended to apply to situations like NDLTD, where we move from a regular DL, through interoperability, to a union DL [84].

Beyond Scenarios, in 5S we have Spaces, Structures, and Streams. **Spaces** clearly can be used to describe the locations of ETD collections around the globe. Spaces also can describe the 2D or 3D interfaces facilitating interaction with DL systems [6] [16] [54] [76].

**Structures** cover all types of organizations, including data structures and databases. Classification of collections based on policies is a simple type of structuring [26]. Documents also can be structured, such as in accordance with the Text Encoding Initiative [12] [102], or through markup encoded using XML [43]. Documents can be classified according to a category system, or using a taxonomy or ontology [81] [86] or other type of knowledge structure [9] [50].

Finally, **Streams** can be used to model the underlying content in DLs. A digital object of any type ultimately is a sequence or stream of bits, though it may be easier to think of ETDs as strings of bytes or characters or words or sentences or pixels or images. With regard to multimedia content, the notion of a stream usually is

quite clear, such as when we think of an audio or video track. Streams also can be used to describe flows, such as of ETDs from students to universities to the global research community. They can be used to describe the flow of work through DLs [4]. When user profiles are consulted and users are alerted to new works through routing or filtering systems [62], we also have a type of stream processing.

## CONCLUSIONS

Since 1987 there has been movement toward open access to the vast literature of graduate research, which includes reports, theses, and dissertations. Global efforts in a broad range of ETD initiatives have benefited from coordination by NDLTD. Making ETDs freely available has clear benefit to student authors, since their works become much more widely read, and they become much more visible in the research community. Likewise, open access to ETDs is of help to universities, since it increases the awareness of their research activities around the globe.

ETD initiatives have positive influence on other open access efforts since students who have prepared ETDs have learned about digital libraries, and have made a constructive contribution to open access through their authoring and submission activities. Further, having engaged in an open access activity, and having learned a bit about the related issues, they may be more likely to be supporters of open access in general.

We have seen how 5S can be used in checklist-form to describe DLs. We have touched on how 5S relates to open access, but a more focused investigation in that regard could be pursued. Of particular interest would be more discussion of Societies and Scenarios, including economic, legal, political, and other social considerations. We encourage such an exploration, building upon the abovementioned involvement of NDLTD and others in worldwide ETD initiatives.

## ACKNOWLEDGEMENTS

## REFERENCES

1. ADOBE. *Adobe Electronic Thesis and Dissertation (ETD) Initiative.* 2006. Disponível em: <http://www.adobe.com/education/etd/>. Acesso em: 2006.

2. ALLARD, S. Communication channels and the adoption of digital libraries for electronic theses and dissertations. In: ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL'04), 2004. *Electronic proceedings…* p. 381. Disponível em: <http://doi.ieeecomputersociety.org/10.1109/JCDL.2004.1336160>. Acesso em: 2006.

3. ANAN, H. *et al.* Preservation and transition of NCSTRL using an OAI-based architecture. In: JCDL'2002, SECOND ACM / IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, 2., 2002, Portland. *Proceedings…* Marchionini: ACM, 2002.

4. ANDERSON, K. M. Metis: lightweight, flexible, and web-based workflow services for digital libraries. In: ACM/IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, 3., 2003, Houston, Texas. *Electronic proceedings…* Disponível em: <http://portal.acm.org/citation.cfm?id=827152>. Acesso em: 2006.

5. ATKINS, A. *et al. ETD-ms:* an interoperability metadata standard for electronic theses and dissertations, 2001. Disponível em: <http://www.ndltd.org/standards/metadata/current.html>. Acesso em: 2006.

6. BAYRAKTAR, M. *et al.* A web art gallery. In: DIGITAL LIBRARIES '98: THE THIRD ACM CONF. ON DIGITAL LIBRARIES, 3., 1998, Pittsburgh. *Electronic proceedings…* p. 277-278. Disponível em: <http://www.acm.org/pubs/articles/ proceedings/dl/276675/p277-bayraktar/p277-bayraktar.pdf>. Acesso em: 2006.

7. BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, v. 284, n. 5, p. 34-43, May 17, 2001.

8. BIENZ, T.; COHN, R.; ADOBESYSTEMSINC. *Portable document format reference manual.* Boston, MA: Addison-Wesley, 1993.

9. BINDING, C.; TUDHOPE, D. KOS at your service: programmatic access to knowledge organisation systems. *Journal of Digital Information*, v. 4, n. 4, 2004. Disponível em: <http://jodi.tamu.edu/Articles/v04/i04/Binding/>. Acesso em: 2006.

10. BORBINHA, J. L. *et al.* A Digital library for a virtual organization. In: ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES (HICSS), 31., 1998, Hawaii. *Proceedings…* S.l.:[s.n.], 1998. V. 2. p. 121.

11. BORGMAN, C.; FOX, E. A. In: JOINT CONFERENCE ON DIGITAL LIBRARIES JCDL'2001, 2001, Roanoke, VA. *Electronic proceedings…* New York: ACM Press, 2001. Disponível em: <http://www.jcdl.org>. Acesso em: 2006.

12. BURNARD, L. Text encoding for interchange: a new consortium. *Ariadne*, n. 24, 2000. Disponível em: <http://www.ariadne.ac.uk/issue24/tei/>. Acesso em: 2006.

13. CALADO, P. P. The web-dl environment for building digital libraries from the web. In: JCDL'2003, JOINT ACM / IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, 3., 2003, Houston. *Proceedings…* S.l.:[s.n.], 2003. p. 346 - 357.

14. CALIS. *NDLTD union catalog mirror site in China.* 2004. Disponível em: <http://ndltd.calis.edu.cn>. Acesso em: 2006.

15. CHIARIGLIONE, L. The MPEG home page. In: ISO/IEC JTC1/SC29 WG11 Moving Picture Experts Group. Torino, Italy: CSELT, 2001. Disponível em: <http://www.chiariglione.org/mpeg/>. Acesso em: 2006.

16. NEVES, F. A. das; FOX, E. A. A study of user behavior in an immersive virtual environment for digital libraries. In: ACM CONFERENCE ON DIGITAL LIBRARIES: DL '00, 5., 2000, San Antonio. *Proceedings…* New York: ACM Press, 2000. p. 103-111.

17. DUBLIN CORE METADATA INITIATIVE - DCMI. *Homepage.* 2006. Disponível em: <http://www.dublincore.org/>. Acesso em: 2006.

18. D-LIB MAGAZINE, 2006. Disponível em: <http://www.dlib.org>. Acesso em: 2006.

19. DUBLIN-CORE-COMMUNITY. *Dublin Core metadata element set.* Dublin, Ohio: OCLC, 2002. Disponível em: <http://dublincore.org/documents/dces/>. Acesso em: 2006.

20. EATON, J.; FOX, E.; MCMILLAN, G. Results of a survey of Virginia tech graduates whose digital theses and dissertations are accessible worldwide. *Council of Graduate Schools (CGS) Communicator,* v. 33, n. 9, p. 1, 7, 8, Nov. 2000. Disponível em: <http://scholar.lib.vt.edu/theses/EatonCGS2000ETDauthors.pdf>. Acesso em: 2006.

21. _____. *The role of electronic theses and dissertations in graduate education. The Council of Graduate Schools Communicator,* v. 31, n. 1, p. 1, 1998.

22. EUROPEAN CONFERENCE ON RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES - ECDL. *Towards the european digital library.* Sept. 2006. Disponível em: <http://www.ecdl2006.org/>. Acesso em: 2006.

23. ELSEVIER. *Scirus.* 2005. Disponível em: <http://www.scirus.com/>. Acesso em: 2006.

24. FOLLETT_SOFTWARE. *Understanding MARC bibliographic: machine-readable cataloging.* Washington, D.C.: Library of Congress, 2000. Disponível em: <http://lcweb.loc.gov/marc/umb/>. Acesso em: 2006.

25. FOX, E.; MCMILLAN, G.; SULEMAN, H. *NDLTD:* Networked Digital Library of Theses and Dissertations. 1997. Disponível em: <http://www.ndltd.org>. Acesso em: 2006.

26. _____. Networked Digital Library of Theses and Dissertations. In: ANDREWS, J.; LAW, D. (Ed.). *Digital libraries:* policy, planning, and practice. England: Ashgate Publishing, 2004. p. 167-187.

27. _____ *et al.* Digital libraries. In: PRACTICAL Handbook of Internet Computing. S.l.]: M. Singh, 2004. Cap. 4.

28. _____. Digital libraries ("hot topics"). *IEEE Computer,* v. 26, n. 11, p. 79-81, 1993.

29. _____. *Sourcebook on digital libraries:* report for the National Science Foundation. Blacksburg, VA: Dept. of Computer Science, Virginia Tech, 1993. (Technical Report, TR-93-35). Disponível em: <http://fox.cs.vt.edu/pub/DigitalLibrary/>. Acesso em: 2006.

30. _____. The 5S framework for digital libraries and two case studies: NDLTD and CSTC. In: NIT'99, 1999, Taipei, Taiwan. Electronic *proceedings…* Disponível em: <http://www.ndltd.org/talks/nit99.ppt>. Acesso em: 2006.

31. _____. The digital libraries initiative: update and discussion: guest editor's introduction to special section. *Bulletin of the American Society of Information Science,* v. 26, n. 1, p. 7-11, 1999.

32. _____. From theory to practice in digital libraries: 5S and educational applications (NDLTD, CSTC). In: NSF-CONACYT-ISTEC WORKSHOP ON DIGITAL LIBRARIES, 1999, Albuquerque. *Electronic proceedings…* Disponível em: <http://www.ndltd.org/talks/nm990708.ppt>. Acesso em: 2006.

33. _____. Update on the networked digital library of theses and dissertations. In: ANNUAL CLINIC ON LIBRARY APPLICATIONS OF DATA PROCESSING, 35., 1998, Illinois. *Papers…* Illinois: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 2000.

34. _____ *et al.* National Digital Library of Theses and Dissertations: a scalable and sustainable approach to unlock university resources. *D-Lib Magazine,* v. 2, n. 8, Sept. 1996. Disponível em: <http://www.dlib.org/dlib/september96/theses/09fox.html>. Acesso em: 2006.

35. _____ *et al.* Networked digital library of theses and dissertations: an international effort unlocking university resources. *D-Lib Magazine,* v. 3, n. 9, Sept. 1997. Disponível em: <http://www.dlib.org/dlib/september97/theses/09fox.html>. Acesso em: 2006.

36. _____ *et al. Electronic theses and dissertations:* a sourcebook for educators, students, and librarians. New York: Marcel Dekker, 2004.

37. _____ *et al.* The networked digital library of theses and dissertations: changes in the university community. *Journal of Computing in Higher Education,* v. 13, n. 2, p. 102-124, Spring 2002.

38. _____ *et al.* NDLTD: Encouraging International Collaboration in the Academy. *DESIDOC Bulletin of Information Technology (DBIT),* v. 17, n. 6, p. 45-56, 1997. Special issue on digital libraries.

39. ACM INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES (DL'96), 1., 1996, Bethesda, MD. *Proceedings….*New York: ACM, 1996.

40. FOX, E. A.; MCMILLAN, G.; EATON, J. The evolving genre of electronic theses and dissertations. In: DIGITAL DOCUMENTS TRACK OF HICSS-32; ANNUAL HAWAII INT. CONF. ON SYSTEMS SCIENCES (HICSS), 32., 1999, Maui, HI. *Electronic proceedings…* Disponível em: <http://scholar.lib.vt.edu/theses/presentations/Hawaii/ETDgenreALL.pdf>. Acesso em: 2006.

41. _____; URS, S. Digital libraries. *Annual Review of Information Science and Technology (ARIST),* v. 36, B, p. 503-589, 2002.

42. GOLDFARB, C. F. *The SGML handbook.* Oxford: Oxford University Press, 1990.

43. _____; PRESCOD, P. *The XML handbook.* USA: Prentice-Hall PTR, 1999. Disponível em: <http://www.phptr.com/bookstore/product.asp?isbn=0130147141&rl=1>. Acesso em: 2006.

44. GONCALVES, M. A. e. *Streams, structures, spaces, scenarios, and societies (5S): a formal digital library framework and its applications.* 2004. Dissertation (Doctoral in Computer Science)- Virginia Tech, Blacksburg, VA, 2004. Disponível em: <http://scholar.lib.vt.edu/theses/available/etd-12052004-135923/>. Acesso em: 2006.

45. _____ *et al.* Streams, structures, spaces, scenarios, societies (5S): a formal model for digital libraries. *ACM Transactions on Information Systems,* v. 22, n. 2, p. 270-312, Apr. 2004. Disponível em: <http://doi.acm.org/10.1145/984321.984325>. Acesso em: 2006.

46. _____ *et al.* Towards a quality model for digital libraries. In: DELOS WORKSHOP ON THE EVALUATION OF DIGITAL LIBRARIES, 2004, Italy. *Proceedings…* Italy: University of Padua, 2004.

47. _____ *et al.* An XML log standard and tool for digital library logging analysis. In: RESEARCH AND ADVANCED TECHNOLOGY

FOR DIGITAL LIBRARIES, EUROPEAN CONFERENCE, ECDL 2002, 6., 2002, Rome, Italy. *Electronic proceedings...* S.l.]: Springer, 2002. p. 129-143. Disponível em: <http://www.springerlink.com/index/6D5505P4WHKN4MVL>. Acesso em: 2006.

48. _____; ZHOU, Y.; FOX, E. A. Providing extended services and resources to the NDLTD community. In: ETD'2002, 2002, Provo, Utah. *Proceedings...* Provo, Utah: NDLTD, 2002.

49. HARRISON, D. ETD implementation in six months: five principles. In: ETD 2004: DISTRIBUTING KNOWLEDGE WORLDWIDE THROUGH BETTER SCHOLARLY COMMUNICATION, 2004, Lexington. *Electronic proceedings...* Disponível em: <http://www.uky.edu/ETD/ETD2004/abstract2.html#devine>. Acesso em: 2006.

50. HODGE, G. *Systems of knowledge organization for digital libraries:* beyond traditional authority files. Washington, D.C.: CLIR, 2000. Disponível em: <http://www.clir.org/pubs/abstract/pub91abst.html>. Acesso em: 2006.

51. INTERNATIONAL CONFERENCE ON ASIAN DIGITAL LIBRARIES -ICADL. *Homepage.* 2006. Disponível em: <http://www.icadl.org/>. Acesso em: 2006.

52. INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES - IJDL. *Homepage.* 2006. Disponível em: <http://www.ijdl.org/>. Acesso em: 2006.

53. JPEG COMMITTEE. JPEG JBIG. In: HOME site of the JPEG and JBIG committees. Crowborough, England: Elysium Ltd, 2001. Disponível em: <http://www.jpeg.org/>. Acesso em: 2006.

54. KENGERI, R. *et al.* Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. *International Journal on Digital Libraries,* v. 2, n. 2/3, p. 157-169, 1999. Disponível em: <http://link.springer.de/link/service/journals/00799/bibs/9002002/90020157.htm>. Acesso em: 2006.

55. KIM, S.; YANG, S.; FOX, E. A. Supply and semand analysis in NDLTD based on patron specialty and contents statistics. In: INTERNATIONAL SYMPOSIUM ON ELECTRONIC THESES AND DISSERTATIONS (ETD'2006), 9., 2006, Quebec, Canada. *Proceedings...* Quebec: NDLTD, 2006.

56. KLETNIEKS, C. *Resources for developers of ETD databases:* ETD-db, 2005. Disponível em: <http://scholar.lib.vt.edu/ETD-db/developer/>. Acesso em: 2006.

57. KRAEMER, B.; ALLARD, S. *ETD 2004:* distributing knowledge worldwide through better scholarly communication. Lexington: University of Kentucky, 2004.

58. LAGOZE, C. *The open archives initiative protocol for metadata harvesting - Version 2.0, Open Archives Initiative.* 2002. Disponível em: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>. Acesso em: 2006.

59. LEFURGY, W. G. PDF/A: developing a file format for long-term preservation. *RLG DigiNews,* v. 7, n. 6, Dec. 15, 2003. Disponível em: <http://www.rlg.org/preserv/diginews/v7_n6_feature1.html>. Acesso em: 2006.

60. LICKLIDER, J. C. R. *Libraries of the future.* Cambridge, MA: MIT Press, 1965.

61. LIU, X. *et al.* Arc: an OAI service provider for cross-archive searching. In: ACM/IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL'01), 1., 2001, New York. *Electronic proceedings...* New York: ACM, 2001, p. 65-66. Disponível em: <http://doi.ieeecomputersociety.org/10.1109/JCDL.2001.10000>. Acesso em: 2006.

62. LOEB, S.; TERRY, D. Information filtering. *Communications of the ACM,* v. 35, n. 12, p. 26-28, 1992. Disponível em: <http://www.acm.org/pubs/citations/journals/cacm/1992-35-12/p26-loeb/>. Acesso em: 2006.

63. LUO, M.; FOX, E. A. ETD search services. In: ETD 2004: DISTRIBUTING KNOWLEDGE WORLDWIDE THROUGH BETTER SCHOLARLY COMMUNICATION, 1., 2004, Lexington, KY. *Proceedings...* S.l.:[s.n.], 2004.

64. LYNCH, C. A. The Z39.50 information retrieval standard part I: a strategic view of its past, present and future. *D-Lib Magazine,* v. 3, n. 4, 1997. Disponível em: *<http://www.dlib.org/dlib/april97/04lynch.html>. Acesso em: 2006.*

65. MCCOWN, F. Search engine coverage of the OAI-PMH corpus. *IEEE Internet Computing,* v. 10, n. 2, p. 66-73, Mar./Apr. 2006. Disponível em: <http://doi.ieeecomputersociety.org/10.1109/MIC.2006.41>. Acesso em: 2006.

66. Mcmillan, G. What to expect from ETDs: if you build it, they will use it. In: DIGITAL LIBRARY SYMPOSIUM, 1999, Cleveland, OH. *Electronic proceedings...* Disponível em: <http://scholar.lib.vt.edu/staff/gailmac/presentations/CWRUGMc1.pdf>. Acesso em: 2006.

67. _____. _____: library issues and responsibilities. In: DIGITAL LIBRARY SYMPOSIUM, 1999, Cleveland, OH. *Electronic proceedings...* Disponível em: <http://scholar.lib.vt.edu/staff/gailmac/presentations/CWRUGMc2.pdf>. Acesso em: 2006.

68. _____; PETERS, L. ETDs: practical, operational, and technical issues for universities implementing electronic theses and dissertations. In: CONFERENCE ON PRESERVATION AND ACCESS FOR ELECTRONIC COLLEGE AND UNIVERSITY RECORDS, 1999, Arizona. *Proceedings...* Arizona:[s.n.], 1999.

69. MEHRINGER, A. The Association of American Publishers' Electronic Manuscript Project: a status report. *Bulletin Medical Library Association,* v. 74, n. 1, p. 27-30, 1986. Disponível em: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=227768. Acesso em: 2006.

70. MILOSAVLJEVIC, B. Models for extensible multimedia document retrieval. In: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA SOFTWARE ENGINEERING (ISMSE'04), 6., 2004, Florida. *Electronic proceedings...* p. 218-221. Disponível em: <http://doi.ieeecomputersociety.org/10.1109/MMSE.2004.52>. Acesso em: 2006.

71. MIT. *DSpace:* durable digital depository. 2003. Disponível em: <http://dspace.org>. Acesso em: 2006.

72. MOXLEY, J. M.; MASIELLO, D.; FOX, E. *ETD guide:* the guide for electronic theses and dissertations. 2002. Disponível em: <www.etdguide.org>. Acesso em: 2006.

73. MURTHY, U.; FOX, E. *The superimposed information project at Virginia tech - homepage.* 2006. Disponível em: <http://si.dlib.vt.edu/>. Acesso em: 2006.

74. PAST CONFERENCES, SYMPOSIA AND MEETINGS NDLTD, 2005, Chicago. *Electronic proceedings...* Disponível em: <http://www.ndltd.org/meetings/index.en.html>. Acesso em: 2006.

75. ONLINE COMPUTER LIBRARY CENTER - OCLC. *XTCat NDLTD Union Catalog.* 2004. Disponível em: <http://alcme.oclc.org/ndltd/index.html>. Acesso em: 2006.

76. PHANOURIOU, C. *et al.* A digital library for authors: recent progress of the networked digital library of theses and dissertations. In: ACM CONFERENCE ON DIGITAL LIBRARIES (DL '99), 4., 1999, Berkeley, CA. *Electronic proceedings...* New York: ACM Press, 1999. p. 20-27. Disponível em: <http://www.acm.org/pubs/citations/proceedings/dl/313238/p20-phanouriou/>. Acesso em: 2006.

77. PLATHE, A. *Workshop on an international project of electronic dissemination of theses and dissertations.* Paris: UNESCO, 1999. Disponível em: <http://portal.unesco.org/ci/en/files/3516/10289070824Summary_report.rtf/Summary+report.rtf>. Acesso em: 2006.

78. POWELL, A. *et al. Guidelines for using resource identifiers in Dublin Core metadata.* 2006. Disponível em: <http://dublincore.org/architecturewiki/ResourceIdentifierGuidelines>. Acesso em: 2006.

79. _____; FOX, E. A. Multilingual federated searching across heterogeneous collections. *D-Lib Magazine*, v. 4, n. 8, Sept. 1998. Disponível em: <http://www.dlib.org/dlib/september98/powell/09powell.html>. Acesso em: 2006.

80. POZO, D. P. V. d. *et al.* Modelagem de bibliotecas digitais usando a Abordagem 5S: um estudo de caso. In: SIMPÓSIO BRASILEIRO DE BANCOS DE DADOS, 19., 2004, Brasília, DF. *Anais…* Brasília:[s.n.], 2004.

81. PRETSCHNER, A.; GAUCH, S. Ontology based personalized search. In: IEEE INTL. CONF. ON TOOLS WITH ARTIFICIAL INTELLIGENCE (ICTAI'99), 11., 1999, Chicago. *Proceedings…* Chicago:[s.n.], 1999. p. 391-398.

82. RICHARDSON, R.; FOX, E. A. Evaluating concept maps as a cross-language knowledge discovery tool for NDLTD. In: ETD2005, 2005, Sydney. *Proceedings…* Sydney:[s.n.], 2005.

83. _____. Using concept maps as a cross-language resource discovery tool. In: ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL2005), 5., 2005, Denver. *Proceedings…* Denver:[s.n.], 2005. p. 256-257.

84. SHEN, R. *Applying the 5S framework to integrating digital libraries.* 2006. Dissertation (Doctoral)- Virginia Tech, Blacksburg, VA, USA, 2006. Disponível em: <http://scholar.lib.vt.edu/theses/available/etd-04212006-135018/>. Acesso em: 2006.

85. _____ *et al.* Requirements gathering and modeling of domain-specific digital libraries with the 5S Framework: an archaeological case study with ETANA. In: EUROPEAN CONFERENCE ON DIGITAL LIBRARIES, ECDL, 2005, Vienna. *Electronic proceedings…* Viena:[s.n.], 2005. p. 1-12. Disponível em: <http://dx.doi.org/10.1007/11551362_1>. Acesso em: 2006.

86. SHUM, S. B.; MOTTA, E.; DOMINGUE, J. ScholOnto: an ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries*, v. 3, n. 3, p. 237-248, 2000. Disponível em: <http://kmi.open.ac.uk/projects/scholonto/>. Acesso em: 2006.

87. SOUTHWICK, S. B.; SOUTHWICK, R. Learning digital library technology across borders. In: JOINT CONFERENCE ON DIGITAL LIBRARIES, 2003, Texas. *Papers…* Texas: [s.n.], 2003.

88. SULEMAN, H. Enforcing interoperability with the open archives initiative repository examiner. In: ACM/IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, JCDL'2001, 1., 2001, Roanoke, VA. *Electronic proceedings…* New York: ACM, 2001, p. 63-64. Disponível em: <http://doi.ieeecomputersociety.org/10.1109/JCDL.2001.10007>. Acesso em: 2006.

89. _____. *Open digital libraries.* 2002. Dissertation (Doctoral)- Department of Computer Science, Virginia Tech, Blacksburg, 2002. Disponível em: <http://scholar.lib.vt.edu/theses/available/etd-11222002-155624/>. Acesso em: 2006.

90. _____ *et al.* Networked digital library of theses and dissertations: bridging the gaps for global access - part 2: services and research. *D-Lib Magazine*, v. 7, n. 8, Sept. 2001. Disponível em: <http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html>. Acesso em: 2006.

91. _____ *et al.* Networked digital library of theses and dissertations: bridging the gaps for global access - part 1: mission and progress. *D-Lib Magazine*, v. 7, n. 8, Sept. 2001. Disponível em: <http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html>. Acesso em: 2006.

92. _____; Fox, E. The Open Archives Initiative: realizing simple and effective Digital library interoperability. *Journal of Library Administration*, v. 35, n. 1/2, p. 125-145, Nov. 2001. Disponível em: <http://www.husseinsspace.com/publications/jla_2001_article_oai.pdf>. Acesso em: 2006.

93. _____. Leveraging OAI harvesting to disseminate theses. *Library Hi Tech*, v. 21, n. 2, p. 219-227, 2003. Disponível em: <http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&contentId=861370>. Acesso em: 2006.

94. _____. A framework for building open digital libraries. *D-Lib Magazine*, v. 7, n. 12, Dec. 2001. Disponível em: <http://www.dlib.org/dlib/december01/suleman/12suleman.html>. Acesso em: 2006.

95. _____. Towards universal accessibility of ETDs: building the NDLTD union archive. In: ETD'2002, 2002, Provo. *Proceedings…* Utah: NDLTD, 2002.

96. _____. Building quality into a digital library. In: ACM CONFERENCE ON DIGITAL LIBRARIES: DL '00, 5., 2000, San Antonio, TX. *Proceedings…* New York: ACM Press, 2000.

97. VAN DE SOMPEL, H.; LAGOZE, C. Open Archives Initiative. 2000. Disponível em: <http://www.openarchives.org>. Acesso em: 2006.

98. _____. The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine*, v. 6, n. 2, Feb.15, 2000. Disponível em: <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>. Acesso em: 2006.

99. VISIONARY TECHNOLOGY LIBRARY SOLUTIONS - VTLS. *Networked digital library of theses and dissertations union catalog.* 2004. Disponível em: <http://zippo.vtls.com/cgi-bin/ndltd/chameleon>. Acesso em: 2006.

100. WEIBEL, S. The state of the Dublin Core Metadata Initiative: April 1999. D-Lib *Magazine*, v. 5, n. 4, 1999. Disponível em: <http://www.dlib.org/dlib/april99/04weibel.html>. Acesso em: 2006.

101. _____ *et al.* RFC 2413, Dublin Core metadata resource discovery. In: IETF 1998, Paris. *Proceedings…* S.l.:[s.n.], 1998. Disponível em: <http://www.ietf.org/rfc/rfc2413.txt>; <http://community.roxen.com/developers/idocs/rfc/rfc2413.html>. Acesso em: 2005.

102. WILLETT, P. *TEL* - Text *Encoding in Libraries*: guidelines for best encoding practices: version 2.1 (March 27, 2006) ed. Washington, D.C.: Digital Library Federation, 2006. Disponível em: <http://www.diglib.org/standards/tei.htm>. Acesso em: 2006.