

Knowledge Transfer: the UNESCO guide for developing countries on electronic theses and dissertations

Susanne Dobratz and Peter Schirmbacher

<dobratz,schirmbacher>@rz.hu-berlin.de

Humboldt-University Berlin Computer and Media Center Unter den Linden 6 10099 Berlin; Germany

Introduction

The UNESCO guide on electronic theses and dissertations aims to develop standards and best practices for the creation/enhancement of electronic theses and dissertations (ETD) projects all over the world, but particularly in developing countries. The aim is to provide an online and a CD-ROM version in English, French and Spanish by April 30th, 2001.

ETDs initiatives are operating all over the world, e.g. NDLTD (<http://www.ndltd.org>), the project Cyberthèses (<http://www.cybertheses.org>), Dissertation Online (<http://www.dissonline.de>). Those projects have been recognized as a cost-effective and self-sustainable mechanism for modernizing IT in higher education institutions. ETD projects involve the joint participation of students, researchers, faculty, staff, administrators and librarians, as well as system and network administrators. They train students, faculty and administrative staff in libraries and media center to cope with multimedia issues. Within this several projects a number of institutions worldwide have already established guidelines for the creation and dissemination of ETDs. The partners vision, and the objective pursued by this guide is the conception and creation of an inter-related system for the distribution of theses on an international level. Researchers and students all over the world now commonly use on-line documents and resources. Far from commercial gains the real needs of students and researchers all over the world lies in the widest possible diffusion and use of research results via free access to theses through a single interface, and without institutional or geographical borders or barriers. Many projects for publishing and distributing theses are already under way around the world. The results are conclusive. Hundreds of thousands of students and researchers have already distributed their thesis, or read and used those of their colleagues. The tendency is clear. With the interest raised by the diffusion of theses on the Web, many institutions are conceiving and implementing these projects.

To collect all the experiences and to derive internationally applicable standards and best practices the UNESCO funds the preparation of a guide presenting these standards and best practices.

Overview of Participating Parties

Initiatives to establish and develop an "Electronic Publishing Culture" and to establish a scholarly electronic publishing are working around the world. The largest of these is the Networked Digital Library of Theses and Dissertations (NDLTD)¹ lead by Prof. Edward Fox from the Virginia Polytechnic Institute and State University (Virginia Tech).

"The concept of electronic theses and dissertations (ETDs) was first openly discussed at a 1987 meeting in Ann Arbor arranged by UMI, and attended by representatives of Virginia Tech (Ed Fox from

¹ <http://www.ndltd.org>

Computer Science and Susan Bright from the Computing Center), University of Michigan, SoftQuad, and ArborText. As followup, Virginia Tech funded development of the first SGML Document Type Definition (DTD) for this purpose, by Yuri Rubinski of SoftQuad. In 1993, at the inception of the Monticello Electronic Library Project, supported by SURA and SOLINET, Professor Edward Fox of Virginia Tech became Co-Chair of its Working Group on Theses, Technical Reports and Dissertations. In 1994 SURA funded a workshop at Virginia Tech to develop plans for electronic theses and dissertations (ETDs), selecting Adobe's Portable Document Format (PDF) and the Standard Generalized Markup Language (SGML) for representation and archiving. To help implement these plans, SURA has funded a research, development, and dissemination effort based at Virginia Tech for 1996." (from the description of the WWW-page of NDLTD)

The main goals of the NDLTD initiative are:

- to enable universities to set up digital libraries, by collecting, cataloguing, archiving, and providing access to electronic theses and dissertations worldwide,
- therefore it is essential that universities discover their potential of their intellectual property and productions, and that universities learn to use and share them,
- to improve university education by effectively sharing technology and knowledge, because progress in science is speeding up. Universities can only keep up with this progress if they use graduate research results and make them more readily and more completely available.
- to enable students to use new technologies and learn about electronic publishing and digital libraries, getting used to new media and technologies for performing their research and spreading their results.

Virginia Tech agreed to finance further development in 1991. Since 1992 Virginia Tech has worked with the Coalition for Networked Information (CNI), the Council of Graduate Schools (CGS), UMI and other interested organizations, helping run a series of design and discussion meetings. Additionally, the University Library's Scholarly Communications Project developed the procedures and systems for processing, archiving, and providing public access to Virginia Tech's graduate research works. So the locally started project at Virginia Tech began to spread out throughout the USA and the whole world, becoming the Networked Digital Library of Theses and Dissertations.

A similar initiative covering the frenchspeaking world is the Cybertheses.org Project, which was originally started by the University of Montreal and the University of Lyon 2. In 1998, the Université de Montréal, along with Université Lumière Lyon2 and Université Senghor, obtained a grant from the Fonds Francophone des infrastructures (FFI). A transfer of expertise toward Lyon2 thus occurred in 1999 and our collaboration resulted in the Cyberthèses programme (<http://www.cybertheses.org>). Thanks to a second FFI grant, these same institutions will soon undertake the production of documentation and of various pedagogical tools in order to hold training workshops in 2001.

Initiatives in Australia, India and South-America followed. The

Australian Digital Library project adapted an early version of the software from Virginia Tech support both university and national level ETD programs.

For India, the University of Mysore, VIDYANIDHI, plans that Mysore University should evolve into a National Centre for ETDs. The Centre at Mysore will eventually develop guidelines for all the issues related to electronic theses and dissertations. The University of Mysore has to deal with a lot of India specific problems, such as typical problems of developing countries, especially: multi-language and multi-script requirements, other diversities in ETD content; developing search interfaces in regional languages; issues relating to metadata in the languages and scripts of the item.

For Latin America ISTEAC is engaged in a number of projects throughout Latin America related to libraries, and is now preparing a small booklet on digital libraries for the region, supported in part by the Organization of American States. ISTEAC will coordinate efforts on the Guide related to the Spanish language, with assistance from other colleagues such as in Mexico and Spain. ISTEAC will translate the Guide into Spanish and promote the concept in Latin America and Spanish-speaking countries.

In Europe we have a quite different situation. Here huge initiatives are barely to be found. Electronic Publishing of Theses and Dissertations is mostly been seen as the own duty of every single university. So the German initiative "Dissertation Online" is one of the most prominent projects within Europe. In 1996 four German learned societies - comprising the fields of chemistry, informatics, mathematics, and physics - signed a formal agreement to collaborate in developing and using digital information and communication technologies (ICT) for their members, scientific authors and readers. Since that day several other societies joined. Within this initiative since 1998, a Germany-wide project "Dissertation Online" (http://www.educat.hu-berlin.de/diss_online) has been up and running. The learned societies involved in the project include chemistry, computer science, education, mathematics, and physics, and five German universities as well as computing centers, libraries and the German National Library (DDB). The original project was directed by Prof. Diepold of Humboldt University, Berlin. Within this project several technologies for processing electronic dissertations at universities were developed. A major focus was on development of extended course materials for authors, library staff and computing centres in order to give instructions to other university libraries on how to build up a document server and a university electronic dissertations service. The Course materials (booklets, software, a small video presentation) for librarians are available via a dissertation portal and information system called DissOnline.de (<http://www.dissonline.de>). At the end of the active sponsoring part of the project, the German National Library (<http://www.ddb.de>) has taken the responsibility for governing further activities in Germany within this field.

The following partners are working on the guide:

- Virginia Polytechnic Institute and State University (Virginia Tech) (USA)
- University of Montreal (Canada)
- Université de Lyon 2 (France)
- Humboldt-Universität zu Berlin (Germany)
- Universidad de Chile (Santiago de Chile) (Chile)
- Australian Digital Theses Project (Australia)
- ISTEAC: Ibero-American Science Technology Education Consortium,
- VIDYANIDHI: Digital Library of Indian Electronic Theses (India)

Contents of the Guide

The Guide will be a next generation version of WWW pages and other content that has evolved over the last four years by groups

around the world that are connected with NDLTD. Some is targeted toward various individuals:

- students (preparing ETDs),
- staff (assisting students, or handling systems and services),
- faculty (guiding students, and discussing issues such as copyright).

At another level, Guide content addresses campus infrastructure:

- campus decisions and plans regarding working with ETDs,
- engineering compromises such as between training/assisting students with standards like XML or incurring higher costs for long-term archiving,
- organizational issues such as what to undertake vs. what to outsource or collaborate on.

Generally spoken, the Guide shall motivate students, faculty, staff, campuses, and nations to establish programmes and projects for the writing and dissemination of scholarly electronic resources. It should give support to developing countries through transferring ideas, principles and technology and make it therefore easy for those countries to establish a program on a campus or a country. The guide shall clarify choices of technology and implementations by showing the implications of each alternative, and which ones work. The Guide covers the following aspects:

- Hard and software requirements
- Network requirements
- Legal requirements
- Data formats
- Metadata
- Archiving
- Access tools
- Model workflow
- Establishment of budget
- Proposals for funding for projects

The first chapter gives an introduction to the genre itself and an overview on ETD initiatives worldwide. It focuses on explaining purpose, goals, objectives of ETD activities and how a help for students to become better prepared as knowledge workers and how to ETDs can improve graduate education, and quality as well as expressiveness of ETDs. The Guide will report on the increasing readership of ETDs, it will show that communicating research results via this way is most successful. It will describe how universities can develop digital library services & infrastructures and which advantages an increase of sharing and collaboration among universities and students will bring for everybody in this context. The first chapter will give a brief overview about the history of ETD activities: 1987-2001 and existing global cooperations and how they can be enlarged in order to give support for developing countries.

The second chapter aims especially towards the support for students. It motivates them to participate within local ETD projects and explains the benefits for students. Those are e.g. to minimise duplication of efforts, to improve visibility and to accelerate work flow and to use and access other scientists electronic resources within a global digital library. It will report on well known sites/resources for ETDs: www.theses.org; www.cybertheses.org/; www.dissonline.org. UML. Within the second chapter the use of retrieval interfaces and mechanisms, the use of classification systems, classification schemes used in different disciplines, the importance of satisfying local requirements will be described aimed at the students point of view. The usage of different word processing systems in order to receive archivable, searchable and internet readable electronic documents will be shown. The preparation of multimedia documents, the handling of the copyright for authors will be explained to students.

The third chapter is particularly aimed towards universities administratives. Here reasons and strategies for archiving electronic theses and dissertations are described and proposals on how to develop an ETD program either locally or on a nationwide or cooperative level (steps, process, collaboration with other institutions, stakeholders) are made. Sample scenarios illustrate different approaches, schedules and workflows to the topic. The third chapter will point out the role of the Graduate School and a graduate program for ETDs, it will consider the role of the library and archives, as well as computing centres. Key concerns and their reasons such as intellectual property rights, relations with publishers will be put into the focus of decision makers at universities. The amount of human resources and expertise needed for an ETD program, as well as sources of funding and an overview of the costs and budgets for such a program are put into the discussion.

The fourth chapter deals with the overall technical issues and will point to existing tools and sites, which are freely available or commercially sold. Here the desired and necessary technical infrastructure (networking, hardware considerations, software considerations) will be examined within different contexts: local, regional, national, global. The production side of ETDs will be covered, regarding the hardware and software needed, how multimedia comes into the game and which scripts and encodings are available. A major part will focus on the problem of document representations conversions, it will give an overview of page description languages, such as Postscript and PDF, on how to handle links, bookmarks, thumbnails, as well as an introduction to markup languages such as SGML/ XML and point to software solutions and DTDs for ETDs,. It will explain how to write directly in XML or how to use conversion strategies, tools and rendering-style sheets. Further technologies for metadata, cross walks between different metadata systems, encryption and watermarking, packaging and post processing of ETDs, checking of ETDs, problems of authentication,

version control, as well as backup strategies and mirroring technologies are described. Systems for the dissemination of ETDs, such as identifying systems (URN, PURL, DOI), metadata models for ETDs, cataloguing systems (MARC, DC, RDF), database and information retrieval issues (Packaged solutions, NCSTRL, Library Automation/OPAC, Harvest usage, other search engines) are all in the scope of this chapter.

Chapter 5 explains which steps have to be taken in order to educate the staff of universities to enable them to cope with these new technologies. It will emphasize the importance of collaboration, local team work and the usage of standards.

The final chapter than looks into the future: how can ETD initiatives like NDLTD be expanded, how can the whole world of scholarly communication be integrated, how can technology changes be managed and interoperability guaranteed. The importance of initiative like the Open Archives Initiative is pointed out. The future vision is given by Edward Fox, the founder of NDLTD:

"In the future, NDLTD plans to offer an increased set of services – not just search but also browsing, annotation, and selective dissemination of information (i.e., routing according to profiles). Searching against millions of works will need to be supported by tools for handling full-text, multimedia content-based matching, query by example, and other approaches. Additional mechanisms for preservation, agreements to enhance performance through mirroring, and flexible handling of works in many of the world's languages will all be needed. Continual evaluation and refinement of services, tailored training and education, and increased sharing and collaboration should help ensure ongoing improvement and eventual fulfillment of the many goals and objectives of ETD programs.

We invite you to learn, participate, and contribute to this cooperative venture!" at <http://www.digitalmediainstitute.org/unesco/>.