

Minutes of the NDLTD Steering Committee Meeting held on 3/21/2001 at the California Institute of Technology (Pasadena) (by Anthony Atkins)

Introductions

Attendees: (Please add/delete/change as appropriate)

David Balatti [david.balatti@nlc-bnc.ca]
Julia Blixrud, ARL
Vivian Bouletreau, U. Lyon, France
Jos Borbinha, National Library of Portugal
Tony Cargnelutti, Australian Digital Theses
Mary Case [marycase@arl.org]
Vinod Chachra [chachrav@vtls.com]
Susanne Dobratz, Humboldt U., Berlin
Enos Esikuri [Eesikuri@worldbank.org]
Donald Fischer [Donald.Fischer@ed.gov]
Edward A. Fox [fox@vt.edu]
Jean-Claude Gu don [jean.claude.guedon@UMontreal.CA]
John Hagen, West Virginia U.
Thom Hickey, OCLC
Delphine Lewis [delphine.lewis@il.proquest.com]
Joan K. Lippincott [joan@cni.org]
Gail McMillan [gailmac@vt.edu]
Charles Myers [crm@adobe.com]
Joseph Moxley [moxley@chumal.cas.usf.edu]
Tom Peters, CIC
Axel Plathe, UNESCO
Peter Schirmbacher [schirmbacher@rz.hu-berlin.de]
Natalia Sokolova, St. Petersburg State Technical University
Shalini Urs, U. of Mysore, India
Eric F. Van de Velde [evdv@library.caltech.edu]

Review of Agenda

NDLTD Status and Statistics

Virginia Tech

Gail McMillan spoke about the current status of the NDLTD and usage statistics from Virginia Tech. There are currently 106 NDLTD members at 90 universities.

Virginia Tech now has over 3,000 ETDs. Usage is growing. Suggested the need to compare online usage with paper usage and statistics from UMI. Gail also pointed out the need for

standardizing our collection and presentation of statistics. Access restrictions are still common, because of the continuing fear of online access as prior publication. Increasing usage from overseas.

Of the NDLTD member universities, 4 require ETDs: Virginia Tech, West Virginia, East Tennessee, and the University of Texas at Austin.

A number of universities have also sent teams to Virginia Tech or hosted Virginia Tech personnel as speakers on the topic of ETDs.

NDLTD

Dr. Fox made a plea for continuing outreach, for making potential members aware of resources. There will be work to try to include Chinese universities in the NDLTD. UNESCO has been active in training and recruiting; the University of Columbia has joined as a result of these meetings. There was activity following the meeting in Catalunya. Hong Kong has increased activity. Australia has a strong national effort. Korea has a growing effort. There is a national effort in India. Singapore has joined. New members in Sweden, Brazil, and Italy, as well as more from Norway.

Growth has been primarily outside of the U.S.

Country/Regional Reports

Russia

Natalia Sokolova spoke about a larger project beginning next year (Electronic Russia), of which ETD efforts will be a part. There are also several contests, including one to develop digital libraries. The requirement to get an award was to get participation of the library. The project at Natalia's university will support a Z39.50 union catalog of traditional and electronic materials. Work will include digitization of great works, including past dissertations.

There is also another state program for improvement of electronic resources in universities.

A question was asked about including Russia in UNESCO efforts. Natalia pointed out that the project was currently limiting its scope so that the project would be more manageable, but is already looking towards collaborative efforts in the European Union and beyond.

Portugal

Jos Borbinha from the National Library of Portugal (also a professor at Lisbon Technical University) spoke about developments in Portugal. Their interest is in acquiring documents for long-term preservation. They cooperate with almost all the national libraries in Europe, looking at metadata issues and union databases. Their system is based on Dienst, and is highly customized. Allows submission in English and Portuguese. Allows submission of ETDs in parts or as a whole. Can export records in MARC format (UNIMARC is used in their library). Their environment has been good for money and staffing, but has had problems dealing with different political issues. Have made progress, are now interested in cooperating with Open Archives efforts. Pointed out that libraries are dealing with issues in grey literature that will eventually be of interest in the commercial world. They also have a few projects adding commercial

publications to digital library. Digitizing of past dissertations is also underway. The software they have developed is provided as a black box that is compliant with previous metadata standards and Open Archives, which other universities can install and customize for their needs. Electronic dissertations are optional in their environment. Talked about the proposed Authority Linking mechanism developed for use with the interoperability metadata standard. There is an upcoming project in their environment that will also need to deal with the issue of authorities, including working groups from DELOS that will produce a working paper.

India

Shalini Urs talked about the current progress in India. Shalini is involved in her university's ETD efforts, but also in the national effort. There is now a government policy in place for all universities to have theses in electronic form and make them available over the network. There is a grant to implement these policies. They are currently studying the workflow at different universities in India. 68 universities have already responded and have given the details of their workflow. From this information, they will develop a workflow model for the larger group. They are also working to deal with dissertations, which typically involves sponsorship by several organizations. Have made alliances with one of these groups to require applicants to create their dissertation electronically. There are now 150 theses in electronic form; the network infrastructure to provide access to the work is not in place yet. Because of the lack of network infrastructure, they are proposing to create a centralized repository for ETDs. There are now 16 universities participating in the pilot study. They are looking at agreement structures to allow additional universities to join.

France

Vivian Bouletreau is working at Lyon to archive theses in TEI format. They started with volunteer students and refining their system. There have been problems with different language sets and presentational elements. As of last year, electronic deposit is mandatory. The need for training in the proper use of word processors was evident; many students use their word processors as electronic typewriters, making the conversion process difficult. In France, there has been a working group on ETDs since 2000, working on standards and formats. Standards have been ironed out. They are now working on drafting national requirements for depositing theses electronically. They are now working on distributing the burden of processing ETDs based on regional centers that contribute to a centralized service and share central technical resources. Are also working to train trainers to disseminate knowledge about creating and processing ETDs.

Germany

Susanne Dobratz reported on the status of the Dissertation Online project in Germany. The funding ended in October of 2000. Several working groups at different universities focused on different aspects of the problem (metadata, document formats, etc.). The funding by DFG has now ended, and the initiative has been opened to other institutions. The German National Library is coordinating further efforts (<http://dissonline.de/>).

There is a new project (DINI) that focuses on more general electronic publishing, and development of library infrastructure.

Several DissOnline partners have their own projects: MathDiss, ProPrint, OAD. Each of these groups will contribute to DissOnline.de. There are guidelines for potential participants, including standards for metadata and document formats, and for securing the integrity and authenticity of documents. Documents will be archived at the German National Library (DDL). Are also considering authority issues.

There is also a spin-off effort to implement Open Archives at German Universities. More information about DINI can be found at <http://www.dini.org>. There is a search portal (THEO) for their Open Archives efforts.

Canada

David Balatti spoke about ETD efforts at the National Library of Canada. Processing 10,000 to 12,000 dissertations, partnership to film dissertations with UMI. Contentville's sales of citation information from dissertations struck a nerve there; made an agreement with UMI to remove citations from Contentville. Held a meeting to discuss ways of dealing with this and other issues related to ETDs. David has now been charged with focusing on ETDs, and bringing together a group of deans, academic faculty, students, and others to look at issues like metadata, formats, licensing.

Australia

Tony Cargnelutti spoke about the status of the Australian Digital Theses project. Each university is responsible for archiving its own theses. Most are paper-based. The national library looks after the national union catalog, but does not have a central role in coordinating ETD efforts, as member institutions are not required to contribute information regarding their theses.

There was a grant to develop a digital repository of metadata regarding theses. Participation in the ADT program and inclusion in the union database is voluntary. The universities have the infrastructure and technical ability to create electronic theses, biggest problem is conservative attitudes among faculty and administrators.

In August of last year, the Council of Australian University Librarians was designated to coordinate the ADT. This opened membership to all Australian universities; there are now 17 members out of the 39 total Australian universities. There are not a lot of ETDs in their collections yet; are more interested in building a good model first. Each member institution uses software based on an earlier version of the software used at Virginia Tech. Member institutions process new theses; no work is being done to digitize existing theses. The metadata collected during the submission process is unqualified. In October of last year, released the second version of their ADT software; enables continual management of ETD files and metadata after their initial creation. The new software allows catalogers to edit metadata, adding local controlled vocabularies. The cataloger also adds a record to the national bibliographic database that contains links to the work within the ADT repository. Their goal is to have all Australian universities signed up by the end of the year. Documents formalizing the role and creation of ETDs will be drafted and brought to legislators later this year. The ADT may expand to include New Zealand and Singapore (in which case it would become the Australasian Digital Theses project). There have been experiments looking at sharing ADT metadata with subject aggregators (portals). Example: AVEL (Australian Virtual Engineering Library), an engineering-centric union database.

Online Services

VTLS

Vinod Chachra presented the work done by VTLS to build a multilingual global union catalog. VTLS will host the catalog, store the original submissions, convert the work to UNICODE, and extract relevant metadata for searching and browsing. Issues including allowing updated information to flow from member institutions to the union catalog. Other issues include deciding on thesauri to standardize terms used between institutions.

The project will be divided into three phases: a pilot phase, a test phase, and the live phase. The pilot phase included contributions from several hand-picked institutions, with language sets including Greek, Korean, and Portuguese.

In the second phase, more universities will be involved. Issues of updating the data over time will be considered, and the overall process will be refined.

In the third phase, participation will be open to all interested parties.

Argued for standardization, talked about problems with mixed-language content. Also talked about limiting submission formats, ideally to one, realistically to a small handful. Need clarification regarding rights management, inclusion of traditional theses and dissertations. Concern was voiced about unnecessary duplication of services already provided by WorldCat, Dissertation Abstracts, etc. Point was made that inclusion of traditional theses may frustrate users looking for the full text of the work; at least in the short term, several people agreed that the project should focus solely on ETDs; group gave general consensus to do this for at least a few years and then revisit the issue.

UMI

Delphine Lewis spoke about her meeting with the CGS regarding their issues. They have turned in a request to `{?}`. They are currently printing electronic theses and dissertations and creating the microfilm from the paper. They are working on producing microfilm directly from the ETD, using e-beam cameras. They are looking at microform and digital archiving formats, and working with CGS to build an XML template for submissions. Dr. Fox brought up the need to coordinate the development of standards with existing work done by the NDLTD.

OCLC

Thom Hickey talked about current work at OCLC relevant to ETDs. They currently have a project to create an Open Archives server containing records extracted from WorldCat. It supports DC and RDF metadata formats, but not MARC. Moving towards developing open source tools like PAIRS, an indexing engine written in Java. PAIRS is the first open source software to implement the SMART retrieval algorithm developed at Cornell. Currently have a tool called SCORPION that suggests Dewey classifications for web pages. Have been developing a tool to suggest LC classifications for content as well. Also developing an RDF toolkit (EOR) that is open source. Dr. Fox pointed out the need to make members aware of these tools through some kind of central registry, similar to the UNESCO clearing house.

UNESCO

Axel Plathe talked about current UNESCO involvement in working with ETDs. Focus on distributing technology and knowledge to developing countries. They are currently working with experts familiar with ETDs to prepare a guide for establishing ETD collections. When complete, the work will be translated into other languages, and will be available both in paper and on the web.

Their office in Montevideo has organized two upcoming ETD training courses in Colombia and Uruguay; another will be organized in Brazil shortly. They also plan to establish a regional network of ETD experts.

The hope is to expand work in the African and Asia-Pacific regions. They would like to establish a plan to coordinate new funding and projects with the work done by the NDLTD to develop cost-effective ETD collections, and to disseminate training.

A question was posed about their work in Africa; there is currently a proposal before UNESCO to fund the creation of a catalog of ETDs.

Shalini Urs talked about the work that has been done on the Guide so far. Developing countries need the expertise to develop ETD collections. There is an outline, which is being fleshed out, and the complete guideline should be available in May.

Guylaine Beaudry from the University of Montreal talked about their involvement in developing the Guide. They are committed to translating the document into French.

Reports

Adobe

Chuck Myers talked about the upcoming version of Acrobat 5. Among the improvements are easier creation of PDF files, the ability to view and edit annotations on-line, better mechanisms for adding digital signatures. They also have done a lot of work to incorporate the structured information provided by an XML document into a PDF file. Their approach uses RDF to encode metadata. Acrobat 5 also will maintain the XML structure as a mapping onto the PDF file. Will preserve the electronic text free from presentation, to avoid problems with hyphenation, for example. Although Adobe makes commercial tools to work with PDF, the PDF standard is a published standard, and there are free tools and libraries being developed by third parties.

Adobe is also becoming more heavily involved in e-books. Standards are developing in that industry. Adobe has also developed a model for delivering encrypted books that can only be viewed on a specific machine.

Also talked about the Collaborative Electronic Notebook group. Also talked about David **Notting(?)**'s work to incorporate structure (including TeX structures) into PDF files.

Argued for PDF as a viable solution for delivering documents, and also a format that preserves some of the structure found in SGML/XML documents. Mentioned a paper on combining SGML and PDF posted on <http://www.impressions.com/>.

ARL/SPARC

Had a publication that Gail served as an advisor for (?). Have continued to promote ETDs as part of the larger transformation of the scholarly communication process. Still not seeing much recognition and interest on the part of faculty. Also mentioned that there were many no votes on the adoption of the Dublin Core as an ISO standard.

CIC

Tom Peters talked about the CIC. The CIC member universities generate about 18% of the dissertations produced in the U.S. Member institutions each have their own processes. Have formed smaller working groups to develop XML schemas that can be used within the CIC. Talked about the CIC as a middle-ground between national/international efforts like the NDLTD and local programs at individual universities.

CNI

Joan Lipincott talked about the Coalition for Networked Information. CNI deals with standards, teaching and learning assessment, and government information issues. The core staff is quite small; most of the work is done within the 200+ member institutions. The CNI has semiannual meetings in the Fall and Spring, with 40 breakout sessions and 2 plenary sessions in 2 days. CNI also has worked to develop best practices for a National Institute for Networking Cultural Heritage. CNI has been involved in and helped to fund the Open Archives Initiative. CNI also has been involved in moving broadband technologies to the end user. NLII has released a first version of their READY project, a checklist for higher education to determine how prepared they are to use network technologies as part of their learning. CNI is also helping to develop facilities (information commons, etc.) where more than one unit in a university are collaborating. The goal is to establish a common web site at Dartmouth where institutions can deposit their information.

Montreal

Guyline Beaudry talked about a report prepared by a consortium of Quebecois institutions. The report details practices for preparing and archiving ETDs. Recommend XML for archiving, PDF for delivery. Also recommend that Quebec join the NDLTD as a consortium. Recommend studies regarding applying new technologies to the evaluation of ETD content by advisors. Mentioned the launch of their new web site <http://www.theses.umontreal.ca>. Also have an electronic journal project and a few others.

Conferences

Proceedings for 2001/2001 ETD conferences

John Hagen talked about the proceedings for the 2000 ETD conference. WVU is now collecting materials that were presented there via the web, and can use the same mechanism to collect materials from this conference. These lists of materials could be presented on the HORIZON web site, and also the web site Library Hi Tech News . Joan Lippincott pointed out the need to make speakers aware of the collection of publications before the conference.

2001 ETD conference

Eric Van de Velde talked about the current conference. So far, 161 people are registered to attend. Proceedings will be coordinated through John Hagen. Handouts for presentations are being collected and linked to the conference web site. Eric and John will coordinate funneling materials to the new database of conference proceedings.

2002 ETD conference

Dr. Fox talked about the need to determine a site and hosting organization for the 2002 ETD conference. Eric Van de Velde talked about the needs to increase the participation of Deans of graduate programs. Dr. Fox talked about having the conference located in the center of the US, or of collocating the meeting with the CNI meeting to reduce travel costs. There was the suggestion of collocating the conference with either the CGS conference or the IFLA conference. It was agreed that if possible, there should be an ETD session in this year's IFLA conference. UT Austin was suggested as a possible site by several people. OCLC was suggested as a possible site by Dr. Fox. It was suggested that it could be coordinated with a regional CGS meeting. There was talk about the expense of holding the conference at a hotel versus a university. Gail suggested holding it at a university near another conference.

Scholarly Studies

There will be studies presented at the conference on topics related to ETDs. Also, there is a Sourcebook being prepared for commercial publication, which will be a completely separate effort from the UNESCO guide.

Membership, publicity, outreach, expansion program

Dr. Fox spoke about the need for further outreach, as Dr. Chachra mentioned in his earlier talk, there are no members from Arab countries. Axel Plathe mentioned the NGO International Association of Universities as a possible group to help in the outreach efforts.

Governance: committees, executive committee, NDLTD Inc., Steering Committee Membership

Dr. Fox talked about the lack of support from CGS, the in-kind donations from vendors, dollar support from SOLINET, and contracts with UNESCO. Talked about creating an NGO funded by member fees. Talked about the difficulty of getting central funding from anyone but members.

Tom Peters talked about whether there is enough work for a full-time coordinator, and whether there are enough members to support a coordinator, and how the addition of new members will increase the work load over time.

If the NDLTD became a program under an existing organization, it would have political connotations.

There was talk about identifying potential sources of income, including conference proceedings, training materials, software. John Hagen talked about encouraging institutions that require and/or have greater experience with ETDs to mentor less experienced institutions. Suzanne Dobratz talked about the need for a secretariat. Dr. Fox talked about potentially having member

fees of \$5,000 a year; Suzanne and Shalini were concerned about smaller universities and developing countries being able to afford these fees. Shalini talked about training being one of the commodities the NDLTD could provide to member institutions.

Tony Cargnelutti talked about establishing regional hubs based on common cultures. The regional hubs would send members to the steering committee. Dr. Fox pointed out that this was the intent of the current model. David Balatti pointed out the problems in attempting to sell text and software generated by the NDLTD, namely that the text is not likely to be a great source of revenue, and a profit motive is counter to the philosophy of making the information freely available. Eric Van de Velde pointed out the cost each university already pays to maintain an ETD collection is already greater than \$5,000, and that a central institution could provide NDLTD in-a-box solutions.

Axel Plathe talked about needing a partner with whom they can establish a contract, rather than having to establish a contract with individual institutions, about establishing the NDLTD as an NGO that could coordinate regional efforts.

Dr. Chachra talked about the need to decide whether we are a conference-oriented organization, or an organization providing technical support.

Joan Lippincott talked about the importance of a visionary leader, and about the importance of establishing a working group to look at the goals of the organization, and the ways in which it could become an NGO.

Vivienne Bouletreau talked about not being comfortable with charging for the materials that are provided for free now.

Thom Hickey mentioned that if Virginia Tech is able to maintain the current level of coordination, there might not be a long-term need for a more established organization. He also mentioned that groups like the W3C have maintained a presence and vision without a central organization.

Jose Borbinha pointed out that the short term technical issues will eventually not be issues, and that there will eventually not be a need for NDLTD (in terms of training, technical support, and outreach).

Chuck Myers talked about the need to identify the goals of a potential group. Goals like establishing standards, templates and guidelines do not require a formal organization. He also talked about organizations where there are levels of membership (voting, etc.) based on the level of contribution.

Natalia Sokolova pointed out that the NDLTD would need to provide quality of services equivalent to OCLC or CNI to justify any cost. She pointed out that Russian Universities would probably be more willing to contribute content than to pay fees.

Guylaine Beaudry talked about the link between membership and services. Talked about the services: organizing a conference, organizing committees, establishing standards. Talked about NDLTD as a 20-year transitional organization, that continues working until there is no longer a need for it.

Dr. Fox talked about the ongoing need for the NDLTD, even if the specific issues change.

Reminded us of the model of Open Archives, where there are repositories of information and a protocol to share the information. Open Archives has standards to include different metadata mappings. Provides a model for creating topical databases.

NDLTD currently collects information in PDF, SGML, XML. Metadata is stored in MARC format, SOIF, etc. We have different workflows in our individual sites.

Need to transform standalone archives into OAI providers. Talked about adding drop-in OAI support to ETD-db software.

Need to prepare an ETD-MS standard that can be encoded in XML and/or RDF.

Union database will accept metadata in different formats.

Authorities are a necessary step; OCLC will present a model for linking to central authority databases.

The union collection is presented as a potential source of added value (de-duping, quality control, other added values).

Current research and existing services include Marian (a system at Virginia Tech), citation database, annotation services, and others.

Centralized services. NDLTD centralized DB would collect information from machine-readable sources. Would build a list of training resources, statistics and logs, preservation policies, IP policies, open source resources, standards and policies.

Standards

There will be a more detailed meeting tomorrow to discuss this topic.

Thom Hickey mentioned the January meeting at OCLC, the draft interoperability standard produced there. Talked about the authority linking mechanism developed at OCLC in support of the standard.

Suzanne Dobratz mentioned the workshop in May of 2000 in Berlin, where different institutions from the US and Europe met and discussed metadata standards.

Archiving and preservation

Julia Blixrud mentioned that several libraries have received Mellon funding to look at preserving electronic data (specifically, electronic journals). The Library of Congress has an archiving project going on, archiving sites like that presidential campaign sites. Working on a plan to develop long-term access to federal government materials. Suggested focusing on archiving issues specific to individual digital objects, but remembering that there are overarching issues between different objects.

Axel pointed out a resolution is being developed by UNESCO that can be used by institutions to promote the preservation of digital documents.

Joan Lippincott talked about the different organizations that are already working on preservation standards, and the need for an organization with a larger view to provide some guidance.

Jose Borbinha talked about his frustration with the lack of progress among some organizations that are intended to deal with preservation.

Submission and Workflow Processing

Tony Atkins talked about developing workflow independent software for managing ETDs in a customizable way while minimizing programming costs.

Wrap-up

Dr. Fox talked about the meeting in September, the group agreed to meet again on September 14.