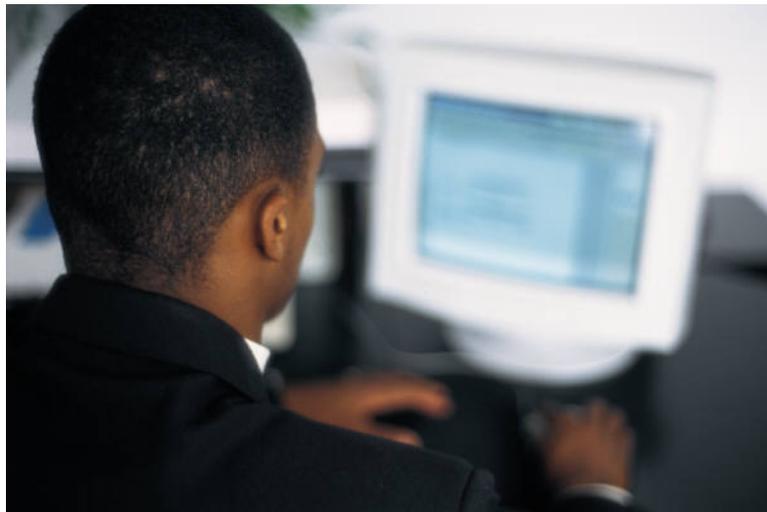


# Feasibility of and Requirements for a South African National Electronic Thesis and Dissertation Project



31 May 2006

Hussein Suleman [hussein@cs.uct.ac.za](mailto:hussein@cs.uct.ac.za)  
Jennifer A. De Beer [jenniferdebeer@gmail.com](mailto:jenniferdebeer@gmail.com)

# 1 EXECUTIVE SUMMARY AND RECOMMENDATIONS

The adoption of Electronic Thesis and Dissertation (ETD) collections has gradually increased at South African higher education institutions over the last 10 years. While historically only metadata was stored and disseminated, a number of institutions are now following the international trend to store full-text and make this accessible online without restriction. The latter mode of Open Access research dissemination has the potential to greatly improve the visibility of the fundamental research output of South African (SA) universities.

Recognising the emergence of this new paradigm, a workshop of university representatives was held in 2003 to introduce and discuss the notion of ETDs. One of the significant outcomes of this workshop was a resolution to set up a South African National ETD project. At approximately the same time, Kaniki [2003] noted that NRF's mandate from government implies that it should host such a national ETD project/portal.

This report outlines the requirements of such a national ETD project, based on information collected from individuals associated with national projects in other countries, as well as a review of the literature on ETD implementations globally.

This study is done at an exciting time when increased momentum is gathering around the adoption of Open Access practices for publication and research dissemination in South Africa's scientific community. Moreover, the latter is expressing itself at the level of the, sometimes tacit, often explicit, endorsement of Open Access scholarly communication from high-level science actors such as the South African Dept. of Science and Technology; the National Research Foundation in collaboration with the Committee on Data for Science and Technology (CODATA), a body of the International Council for Science (ICSU); not forgetting the recent recommendations of the Academy of Science of South Africa (ASSAf). These policy initiatives emphasise the very active role that should be taken by national funding bodies in the digital archiving of scholarly research.

The South African national funding body, the National Research Foundation (NRF), is committed to supporting and playing a major role within the National ETD project, as it is of strategic importance and falls within its mandate of promoting and supporting research, partly through the "...promotion of the provision of an information infrastructure linking research institutions to facilitate cooperation and sharing of research.." and thus through the dissemination and utilization of research results. An early expression of the latter support by the NRF is the very funding of this feasibility and requirements study.

That said, as a concrete way forward, this ETD report recommends the following:

- NRF (Executive Director KM&S) and CHELSA should establish a Reference Group/Steering committee with the general responsibility of compiling the Terms of Reference for the National ETD project and the way forward, and a Technical Committee (tasked with establishing standards, best practices and a portal).
- NRF should establish a national ETD portal with discovery and access services, and possibly archival services.

- NRF should in principle, be amenable to hosting ETD collections for institutions that are not capable of doing this for themselves.
- NRF should host a meeting of representatives of South African universities to discuss and clarify the ownership and rollout of a National ETD Project and determine which institutions need direct hosting support.
- The national ETD portal should be so setup that it can be easily moved to another location should the need ever arise.
- Technical aspects of a national ETD project could be addressed in collaboration with or as a project initially commissioned to a research laboratory with relevant expertise, such as the Advanced Information Management laboratory at the UCT Department of Computer Science – headed by one of the co-authors of this report. This could accelerate the process and ensure that current best practices in the creation of Digital Library systems are operationalised, while also ensuring a strong link between production services and related research and supporting the training of students wherever possible.
- NRF should budget for equipment and staff support for the first year and subsequent years of the project. In-house IT staff may be appropriate for ongoing support after year 1, but initial work may need to be outsourced on a consultative basis.
- NRF might need to consider how its legacy NEXUS bibliographic database of theses and dissertations will be migrated to the proposed national ETD project; and further, how to go about obtaining the associated fulltext, the latter being a necessary requirement for the national project.
- NRF should bear in mind that though the remit of this study has been ETDs, the possibility strongly exists that institutions may want to implement institutional repositories (IRs) hosting the research output of academics and researchers, in addition to graduate research output; and that the NRF should anticipate its role in this regard
- NRF should join the NDLTD and actively participate in the international movement to support ETD collections.

## **2 CONTEXT**

### **2.1 Electronic Theses and Dissertations**

A thesis or dissertation is a document submitted in partial or whole fulfillment of the requirements for a postgraduate or higher degree. These documents have traditionally been prepared on paper, bound in book form and submitted to the university for archiving. More recently, like all other publications, theses and dissertations have been prepared almost exclusively on computer, offering the possibility for electronic storage and dissemination. This move to an Electronic Thesis or Dissertation (ETD) has gradually been adopted by institutions worldwide and has completely replaced older paper-based workflows in many institutions.

ETDs are an example of electronic publishing that is made possible because of a number of core factors:

- Most universities with postgraduate programmes already require theses or dissertations.
- Theses and dissertations are high quality reviewed documents, therefore good indicators of the research done at the institutions.
- Copyright is usually owned by the student and/or the university and its representatives.

ETDs that are produced and archived at an institution will only achieve their full potential in this new information age if they are made accessible on the World Wide Web, typically stored and managed by a digital library software system (such as DSpace).

The transferring and implementation of research results and findings is crucial and a key determinant of the economic performance of the national innovation system of South Africa. The electronic publication of theses and dissertations (ETDs) is therefore imperative for the transfer of knowledge, capacity building as well as cost-effective management of knowledge generation.

If an institution collects and disseminates ETDs online, in addition to or instead of paper-based documents, there are numerous advantages, including:

- Many theses and dissertations are shelved with no efficient way for researchers to locate the information that may be contained in them
- Time and costs involved in procuring copies of these works may often be prohibitive
- Free access to full texts could help scholars by allowing full-text searching to ensure that the documents they order or download actually contain the information they seek

- Electronic publication can make access and distribution faster and less expensive for most scholars
- ETDs can help to make information more readily available to scholars and researchers by allowing quicker and more thorough search capabilities
- ETDs are easily backed up, so the risk of losing information is minimal
- ETDs stored electronically are less likely to be damaged than their print counterparts
- Every postgraduate student will have the experience of authoring and participating in at least one high quality electronic publication process before graduation, thus training a new generation of information-literate graduates.
- Libraries can save drastically on shelf space and management costs [McMillan, 2004] if documents are only stored in electronic format.
- The research conducted by the institution will get higher visibility if the collection of ETDs is accessible online.
- Students and researchers will have access to ETDs of other students, and so avoid re-inventing the wheel vis-à-vis research topics and results.

These advantages highlight an important feature of ETD collections at institutions: the maximum benefit is derived if more institutions participate in such an initiative. If only a handful of institutions host online ETD collections, students at those institutions will only have easy access to a small sample of research conducted elsewhere. However, if every institution in the world hosted its ETDs online, it could potentially improve the quality of research conducted in general. While this noble goal is difficult to achieve on a global scale, some countries have established national ETD projects to create a country-specific research-commons, which then feeds into international coordinating efforts.

This report particularly discusses the requirements for South Africa to establish such a national ETD project and requirements at the institutional level, primarily in terms of technological and human resource aspects. We go on to consider the international scope for collaboration within the ETD sphere, particularly at a technological level.

## 2.2 ETD Projects

ETD Projects are implemented at various levels of granularity, with different requirements at each level.

Institutional ETD projects aim to archive and disseminate the ETDs produced by students, and possibly staff, at the institution. The services provided typically include submission, metadata creation/editing, review, archiving, search, browse and retrieval. These services are usually provided by a software suite such as EPrints<sup>1</sup>, DSpace<sup>2</sup> or the older ETD-db. There are sometimes also defined strategies for location-independent

---

<sup>1</sup> <http://www.eprints.org>

<sup>2</sup> <http://www.dspace.org>

identifiers and preservation. In almost all cases, the collections of ETDs support interoperability protocols such as the OAI-PMH (see Section 4.1).

The number of institutions with ETD projects numbers well over 200 around the world (a full listing of this can be found on the website of the Networked Digital Library of Theses and Dissertations<sup>3</sup>). In addition, some of these represent not individual institutions but regional or national consortia of institutions.

National and regional ETD projects aim to archive and disseminate the ETDs produced by member institutions. The services provided typically include archiving, search, browse and retrieval. In most cases, the software used to provide such services is custom-written, typically obtaining metadata and/or data from member institutions to provide cross-archive services.

Examples of successful national ETD projects include:

- Australasian Digital Theses Programme. This covers institutions in Australia and recently also New Zealand. Similarly to an institutional project, the ADT project republishes its collected metadata using the OAI-PMH [Campbell, 2005]. As of early 2006, ADT had more than 4000 records from 31 active members, at least 12 of whom required every student-authored thesis/dissertation to be submitted in electronic format [Wells, 2005].
- Brazilian National ETD Project (BDTD). 14 Universities in Brazil currently contribute over 12000 records to this union catalogue. The Brazilian effort has concentrated on advocacy and training in addition to setting up metadata standards and a union metadata catalogue [Pavani and Suelli, 2005].
- DiVA. This covers institutions primarily in Sweden but with some participation from other Scandinavian countries. Local archives use a standardised rich XML template to generate full-text and metadata. DiVA operates a portal that provides access to data that is collected and republished from members (using OAI-PMH) [Müller, et al, 2003].

## 2.3 NDLTD

The Networked Digital Library of Theses and Dissertations (NDLTD) was launched in 1996 to support the adoption of archiving of and access to ETDs at institutions around the world [Suleman, et al, 2001]. This is a non-profit organisation that supports efforts at an individual and national level wherever possible. Among its core activities, NDLTD provides the following useful services:

- Setting of international metadata standards
- Collecting and republishing metadata at an international level
- Provision of services and collaboration with international service provision partners (such as Elsevier and Google)

---

<sup>3</sup> <http://www.ndltd.org>

- Annual symposium on ETDs (Sydney in 2005; Quebec City in 2006; Sweden in 2007)

South African institutions were among the first members of NDLTD and continue to be active participants in its activities.

## 2.4 South African ETD Projects

Of the 23 higher education institutions in South Africa, to date seven institutions have implemented ETD projects, viz. Rhodes University's Institution-wide repository; and six ETD repositories at the University of Johannesburg, University of Pretoria, the University of the Free State, University of South Africa, University of the Western Cape, and the University of the Witwatersrand.

The Rhodes eResearch Repository, being an institution-wide repository hosting research output of academics in addition to that of graduate student output, introduces the possibility for institutions to expand their collections from being ETD-centric to institution-wide repositories. Current ETD implementations may evolve to a state of being institutional repositories (IRs). Alternatively, those institutions that have not yet implemented any ETD/IR initiative may decide to initiate their projects as IRs that include ETDs from their very inception. Technically, this does not pose any problems as the sub-collections are uniquely identifiable and separable for the purposes of interoperability with national and international partners (as evidenced by Rhodes' recent participation in NDLTD). From a practical perspective, this marriage between ETD collections and IRs allows institutions to support multiple types of electronic collections using a shared pool of staff and equipment – thereby providing their constituencies with a rich palette of user services. While ETD collections have been the primary focus of this study, the relevance of IRs as a sister technology and information service must be noted.

Regarding the software being used in South African ETD implementations, five repositories<sup>4</sup> are NDLTD members, using the NDLTD VT ETD-db software; Rhodes University uses EPrints software, and one institution, viz. the Western Cape, link directly to the full-text from an alphabetical index Web page [Tise, 2005]. At the time of writing, the University of the Western Cape is in the process of implementing an in-house ETD OAI-compliant software development project [Van der Walt, 2006]. Furthermore, Wits University has migrated from the VT ETD-db software to DSpace, and at the time of writing, the University of Pretoria is in the process of migrating to the use of DSpace.

It is notable that of the repositories listed, the repositories of Rhodes, Johannesburg and Pretoria are registered as Open Access repositories in the OpenDOAR<sup>5</sup> and ROAR<sup>6</sup> registries. Further, where projects are indicated as using the NDLTD ETD-db software, and are hence OAI-compliant and have unrestricted access, it is not entirely clear why these projects are not listed in registries such as ROAR and OpenDOAR. Nevertheless,

---

<sup>4</sup> Universities of Johannesburg, Pretoria, Free State, Witwatersrand, and South Africa (Tise, 2005)

<sup>5</sup> For the OpenDOAR list of South African projects see

<http://www.opendoar.org/doar?func=browse&format=full&cId=177&tId=&cpid=&submit=List>

<sup>6</sup> For a ROAR listing of South African projects see

<http://archives.eprints.org/?action=home&country=za&version=&type=&order=name&submit=Filter>

all repositories are easily discovered using the de facto Google approach to information discovery.

## 2.5 South African ETD/OA Workshops

While South African ETD repositories were being set up in isolation from one another, it was recognised that a concerted coordinated effort would result in better service delivery at both the institutional and national levels. At the Workshop on the Role of Electronic Information Resources in Enhancing and Supporting Research Output in Higher Education Institutions (HEIs), held at the NRF on 30 September 2003, the NRF's role for the access to and utilization of knowledge and more specifically the coordination of electronic thesis and dissertations at a national level were presented by the Executive Director of Knowledge Management and Strategy. [Kaniki & Van der Berg, 2003]

In the same month, a separate workshop was held in South Africa solely to discuss ETD projects, with funding from UNESCO (Ubogu, 2004). This workshop was well attended by representatives from the majority of South African universities, including library directors and staff, research administrators and IT staff. The purpose of the workshop was to provide a motivation for and information on how to set up ETD projects at individual institutions. Since 2003, the number of ETD projects has approximately doubled, partly due to the discussions held at this workshop. One outcome of the workshop was the setting up of a committee to further the process of a national project through the auspices of CHELSA.

Since 2003, three related events have been held around the theme of Open Access. The first was a conference, with invited speakers, to provide information on Open Access and Institutional Repositories for the local community. The events subsequent to the conference were hands-on workshops where participants from local institutions were trained in the use of digital library software (DSpace and Greenstone<sup>7</sup> respectively). These events were relevant to ETDs because the software used for Open Access projects is in most cases the same as that used for ETDs and, as discussed earlier, some institutions (both internationally and in SA) have opted to set up dual-purpose ETD/IR systems, recognising that ETDs are a subset of the information types that are typically covered by collections housed in IRs. These OA events have thus built on the initial ETD impetus and supported the creation of ETD projects through advocacy and staff training.

These events notwithstanding, it has been acknowledged that there is still a lack of coordination among ETD collections within the country, coupled with the near exclusion of smaller institutions (mostly historically disadvantaged institutions) that possess neither the skills nor the resources to set up ETD projects. The national project has been raised as a potential solution to the aforementioned problems.

---

<sup>7</sup> <http://www.greenstone.org>

## 3 THE ROLE OF A NATIONAL PROJECT

### 3.1 Motivation

The establishment of a national project is motivated by factors both external and internal to the National Research Foundation (NRF). The external factors are those originating from initiatives in the Open Access scholarly communications arena at international and national level, which in turn have a direct bearing on the strategic initiatives of the NRF.

One such international initiative is that of the International Council for Science (ICSU) Priority Area Assessment (PAA) Panel on Scientific Data and Information. The panel recommends<sup>8</sup> that ICSU members, interdisciplinary bodies, **science funding bodies** (our emphasis), data and information providers, and users, as key stakeholders, should be involved in the development of the long-term strategic framework to ensure that the full benefits of new data and information technologies and capabilities are extended to scientists throughout the world [ICSU, 2004].

The National Research Foundation in partnership with the South African National Committee for CODATA and the United States National Committee for CODATA (one of ICSU's bodies) organized a workshop<sup>9</sup> with the following objectives:

1. Review the current status of practices for sharing and archiving scientific information resources related to sustainable development in the SADC region
2. Identify and discuss scientific, legal and policy, institutional and economic, and management and technical factors relevant to providing permanent access to digital scientific information resources. Examine different models, and their benefits and shortcomings in the SADC region, drawing on examples of related digital archiving and access regimes.

This workshop resulted in recommendations regarding i.a. mandated access to data and information, and suggests policy interventions on the part of research funders which would take the form of required self-archiving on the part of supported researchers. Most notably, the Executive Summary [CODATA, 2006] recommends:

- 1 Establish and implement policy interventions by research funders (including governments and institutions) that:
  - 1.1 Mandate that scholars make pre-prints and e-prints of their research available via an open access medium.
  - 1.2 Mandate long-term curation of research outputs, both in the interests of the individual researchers who produce the articles, but also in recognition of the shared character of the global research enterprise.
- 2 Promote the value of open-access approaches to the research funding bodies by:
  - 2.1 Involving researchers and managers in describing the real challenges as well as solutions (using available local success stories).

---

<sup>8</sup> [http://www.icsu.org/1\\_icsuinscience/DATA\\_Paa\\_1.html](http://www.icsu.org/1_icsuinscience/DATA_Paa_1.html)

<sup>9</sup> <http://stardata.nrf.ac.za/html/workshopCodataPublications.html>

- 2.2 Establishing training programmes for researchers and for journal funders and producers.
- 3 Create high-quality regional information repository facilities where individual publications, or the output of small subgroups of scientists, can be cost-effectively preserved, and openly available.
  - 3.1 This will support the digitisation of more African material.
  - 3.2 Promote the establishment of open institutional repositories.
  - 3.3 Include national repositories to archive national heritage items and provide quality-control functions such as selection, appraisal and retention. [p.8]

Another initiative, at national level this time but with a bearing on the NRF, is that of the Academy of Science of South Africa (ASSAf). Of particular relevance here is part of Recommendation No. 6 arising from the recently completed SA Journals study of the Academy of Science of South Africa [2006]:

Recommendation No. 6: that the Department of Science and Technology takes responsibility for ensuring that Open Access initiatives are promoted to enhance the visibility of all South African research articles and to make them accessible to the entire international research community. Specifically:

- Online, open access (“Gold route”) versions of South African research journals should be funded in significant part through a per-article charge system...;
- A federation of institutional Open Access repositories, adhering to common standards, should be established (“Green route”), with resources made available to help institutions in the preliminary stage, this virtual repository to be augmented by a central repository for those institutions which are unable to run a sustainable repository;
- National harvesting of South African Open Access repositories should be undertaken as a matter of urgency, preferably by the NRF (our emphasis)

These recent events strongly underscore the need for a national body such as the NRF to promote and provide exposure to the research output of next-generation South African researchers, as expressed through the particular research themes and domains covered in postgraduate research theses and dissertations. Moreover, the National Research Foundation Act, No. 1998 stipulates under the heading "Functions, powers and duties" section 4.(1)(n) promote the provision of an information infrastructure linking research institutions to facilitate co-operation and sharing of research information and knowledge. As such, it is virtually a moot point that NRF needs to host (or commission the hosting of) a national portal and archive of research as reflected by ETDs.

The inception of national ETD projects in other countries (notably, Australia and Brazil) has resulted in a marked increase in the number of institutions with ETD projects. This effect is sorely needed in South Africa, where some institutions have enthusiastically embraced the concept of electronic publishing of theses and dissertations while others have not. NRF’s role would thus implicitly and explicitly include that of advocacy for

ETD projects. In addition, as the nominal guardian of national research, and considering the external factors mentioned earlier, along with current discussions on access to publicly-funded research, NRF may adopt the role of a policy-maker in conjunction with its ETD efforts. NRF is in a prime position to develop and operationalise such policies.

Coordination of ETD projects is another reason for a national authority. While each project can function autonomously, the recent IR workshops have illustrated how sorely South African archivists need technical skills to make the leap to the electronic age. A national ETD project could thus include components for training of staff, development of shared standards and sharing of knowledge and expertise.

## **3.2 Functions**

The first and most obvious function of the NRF, as host of a South African National ETD project, is to maintain a portal where students and researchers may search through the collections of ETDs located at institutions within South Africa. The portal may provide the ability to search through the metadata and/or the full-text of the documents.

A short survey was sent to individuals involved in ETD projects to gather requirements in terms of services for a national ETD project. While there were only a handful of respondents commenting on national projects, they largely agreed that the services should include: hosting of search and retrieval services; harvesting of metadata from member sites; hosting of a union metadata catalogue; standards development; and participation in NDLTD. Most of these services are handled by a portal infrastructure.

Coupled with a portal for access, the National ETD Project can and should maintain an archival copy of the full-text of all documents which would serve as a fixed record of research, independent of institutional changes. This is particularly significant in the climate of institutional transformation and mergers. Moreover, the hosting of an archival copy at national level goes some way to addressing the preservation of these scholarly works.

Where it is anticipated that an institution may not have the resources nor capability to host its own ETD repository, two solutions may be had by way of the NRF hosting such a full-text server on behalf of an institution, or where an institution with an extant ETD repository may agree to host the documents of a fellow institution in need of such a service.

Lastly, the national project may host (online and in-person) fora for discussion of policy issues, collaboration and training of staff. These fora may take the form of mailing lists or workshops, as and when the need arises.

## **4 ARCHITECTURE OF A NATIONAL DISTRIBUTED ETD SYSTEM**

### **4.1 Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)**

To satisfy the national project's role as a central portal for all ETDs produced in the country, it is necessary to gather metadata and/or full-text documents from all institutions into a single catalogue where researchers can locate items of interest.

In 1999, a number of distributed archive communities met to discuss the problems of connecting together remote collections. Federated search, exemplified by the NCSTRL project [Leiner, 1998], did not work where remote sites were operated independently because sites changed their software/hardware or became disinterested over time. Also, the federated search standard of the day, Z39.50 [NISO, 2003], was deemed too complex and difficult to support in a large distributed community. As a result, a simpler, more robust, low-barrier interoperability standard [Lagoze and Van de Sompel, 2001] was devised, based on the principle of periodically exchanging all updated metadata records. This evolved into the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [Lagoze, et al, 2002].

OAI-PMH is a Web-based protocol to transfer XML-encoded metadata from one machine to another, on an incremental basis. Thus, a central metadata repository need only contact each remote site at discrete intervals to request updates since the last batch. The remote sites, that host data collections, are referred to as data providers. The central metadata repository is referred to as a service provider, because of its role in providing cross-archive search/retrieval services. The service provider periodically harvests (obtains) metadata, using the well-defined OAI-PMH, from its data providers and incorporates this metadata into its catalogue.

The OAI-PMH does not provide an explicit mechanism to transfer full-text objects – only metadata. However, it is possible to define a metadata format incorporating pointers to full-text objects that can then be downloaded automatically. Theses Canada uses this approach to obtain full-text documents [Reeves, 2005].

The OAI-PMH is used by every federated national ETD project documented in this study (Brazil's BDTD, Sweden's DIVA, Australasia's ADT, etc.) as it has proven to be a reliable and robust technology for connecting together independent archives.

Figure 1 shows a typical response from an OAI-PMH data provider – in this case for an identification of the archive.

```

<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2006-04-20T13:50:30Z</responseDate>
  <request verb="Identify">http://txspace.tamu.edu/dspace-oai/request</request>
  - <Identify>
    <repositoryName>TxSpace at Texas A&M University Libraries</repositoryName>
    <baseURL>http://txspace.tamu.edu/dspace-oai/request</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>cgreen@lib-gw.tamu.edu</adminEmail>
    <earliestDatestamp>2001-01-01T00:00:00Z</earliestDatestamp>
    <deletedRecord>persistent</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    <compression>gzip</compression>
    <compression>deflate</compression>
  - <description>
    - <toolkit xsi:schemaLocation="http://oai.dlib.vt.edu/OAI/metadata/toolkit http://oai.dlib.vt.edu/OAI/metadata/toolkit.xsd">
      <title>OCLC's OAICat Repository Framework</title>
      - <author>
        <name>Jeffrey A. Young</name>
        <email>jyoung@oclc.org</email>
        <institution>OCLC</institution>
      </author>
      <version>1.5.38</version>
      <toolkitIcon>http://alcm.e.oclc.org/oaicat/oaicat_icon.gif</toolkitIcon>
      <URL>http://www.oclc.org/research/software/oai/cat.shtm</URL>
    </toolkit>
    </description>
  </Identify>
</OAI-PMH>

```

Figure 1. Response from institutional ETD project OAI-PMH interface

## 4.2 Metadata Standards

### 4.2.1 Dublin Core

OAI-PMH requires that every data provider have the ability to express every record in Dublin Core (DC), which is a simple metadata format designed for general resource discovery online [Dublin Core Metadata Initiative, 2004]. ND LTD recommends a specific interpretation of the Dublin Core standard for theses (see ETD-MS section). Figure 2 shows an example of a typical DC record for an ETD.

```

<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <title>Open Digital Libraries</title>
  <creator>Suleman, Hussein</creator>
  <subject>Computer Science</subject>
  <description>Digital Libraries (DLs) are software ...</description>
  <contributor>Roger W. Ehrich</contributor>
  <contributor>Srinidhi Varadarajan</contributor>
  <contributor>James D. Arthur</contributor>
  <publisher>VT</publisher>
  <date>2002-11-26</date>
  <type>text</type>
  <format>application/pdf</format>
  <identifier>
    http://scholar.lib.vt.edu/theses/available/etd-11222002-155624/
  </identifier>
  <source>
    http://scholar.lib.vt.edu/theses/available/etd-11222002-155624/
  </source>
  <language>en</language>
  <rights>unrestricted</rights>
  <rights>I hereby certify that ...</rights>
</oai_dc:dc>

```

Figure 2. Sample Dublin Core record for an ETD

#### 4.2.2 ETD-MS

NDLTD developed and maintains the ETD Metadata Set (ETD-MS) standard in consultation with worldwide partners to augment Dublin Core with additional fields for the granting institution, level of degree, name of degree and discipline [Atkins, et al, 2006]. Figure 3 shows the unique parts of a typical ETDMS record for an ETD.

```

<oai_etdms:thesis
xsi:schemaLocation="http://www.ndltd.org/standards/metadata/etdms/1.0/
http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd">
  ...
  <degree>
    <name>PHD</name>
    <level>doctoral</level>
    <discipline>Computer Science</discipline>
    <grantor>VT</grantor>
  </degree>
</oai_etdms:thesis>

```

Figure 3. Fragment of ETD-MS record for an ETD

### 4.2.3 Country-Specific Metadata

Some national projects (e.g., Brazil, Germany, Sweden) have developed or are developing specific metadata formats to cater for internal needs. These needs include the requirements for legal deposit in the respective national libraries, language-specific issues, national identifier systems and existing best practices. In most cases, these metadata sets are simply supersets of DC or ETD-MS.

## 4.3 Distributed Architecture

Figure 4 illustrates the distributed architecture of a typical seamless ETD solution for South Africa, based on current standards and best practices elsewhere as well as the requirements of the local situation. In the diagram, all arrows represent harvesting of metadata using the OAI-PMH.

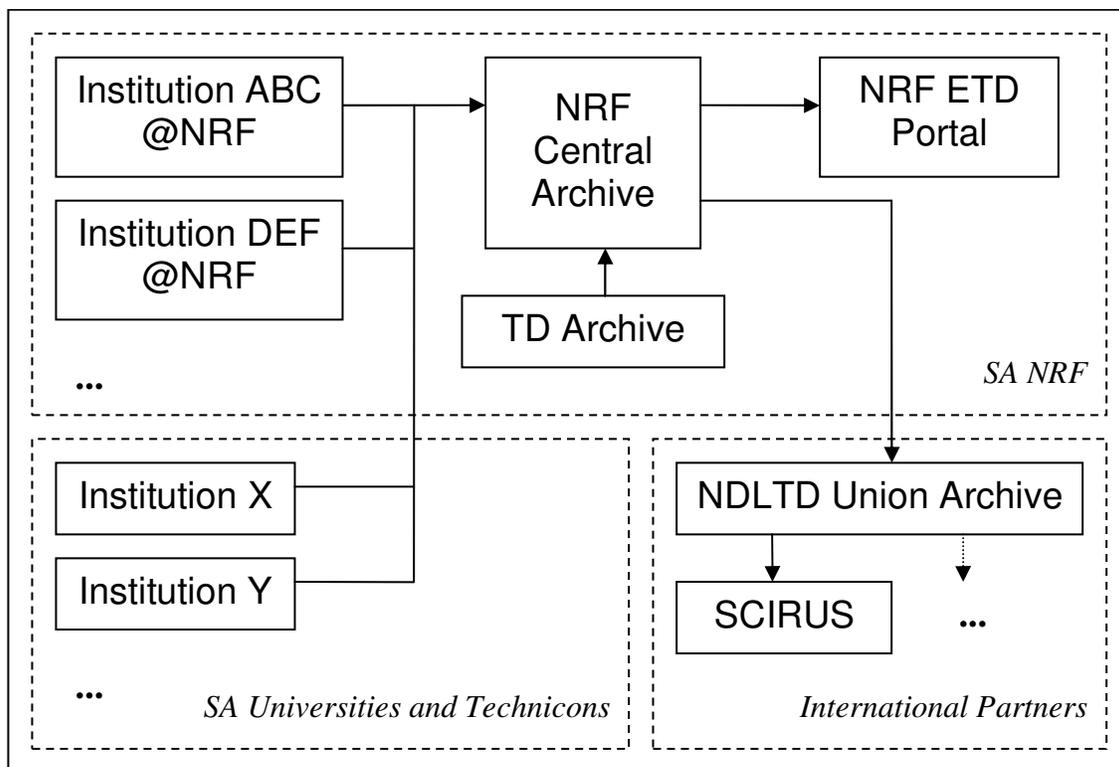


Figure 4. Distributed architecture of a seamless ETD system

NRF would operate a central metadata and digital object archive, which obtains data updates from numerous OAI-PMH-compliant sources on a regular basis. The ideal source is a university archive for ETDs or that contains ETDs as a clearly-defined subset – labeled Institution X and Institution Y on the diagram. If an institution does not possess the capacity to host its own ETD system, NRF can host the collection as an independent entity, thus making it trivial to move to the owning institution at a future date – labeled Institution ABC and Institution DEF on the diagram.

Finally, the central archive may also include currently-existing metadata records for theses and dissertations not in electronic format (TDs). Specifically, the current NEXUS database can be connected into this system easily by exporting the data and creating a

simple fixed collection OAI-PMH archive using the Static Gateway Specification (Van de Sompel et al., 2004] and software distributed by the OAI.

The central archive would serve as an OAI-PMH data provider. NRF's portal can then harvest records from it in order to support a search and retrieval interface. Similarly, NDLTD can harvest records to incorporate South African ETD/TD metadata into its international catalogue (and, indirectly, entities such as Scirus).

It should be noted that the NRF is willing to take the role as a hosting organization for institutions reflected as ABC and DEF above, noting especially that it will be a challenge to work with different institutions and a responsibility to ensure proper training of all parties. Nevertheless, the NRF has indicated that it is keen to adopt this role.

## **5 REQUIREMENTS FOR A NATIONAL PROJECT**

A short survey – selected responses for which can be found in Appendix B – was sent to individuals involved in ETD projects to get estimates for the staff and equipment requirements of a national ETD project. In addition, where applicable, this information was derived from publications by existing national projects elsewhere. It is expected that the NRF (and CHELSA) will engage in an in-depth costing exercise of the proposed national project following the submission of this report.

### **5.1 Personnel and Training**

#### **5.1.1 Personnel**

The survey results indicate that the number of staff required to manage institutional ETD collections varied between 1 and 2, and staff fell into the following categories:

- computer technician / programmer
- librarian
- administrative clerk
- student assistant
- manager

In addition, a multidisciplinary committee is usually responsible for the management of the ETD project. Finally, additional staff support, especially from the IT division, is typically required when starting a new project or performing advanced maintenance activities (such as migrating to new equipment).

The distinction between staffing for institutional and national projects is fuzzy, as a lead institution is usually responsible for the national project's technical requirements (e.g., IBICT in Brazil, UNSW in Australia). As such, respondents indicated negligible (up to 10%) additional staff for maintenance of a union archive and portal.

In the case of a South African national project, administrative clerks will not be necessary at NRF as these are the individuals who submit ETDs and/or review submissions at the institutions. However, those institutions which do not have institutional capacity might

do well to have a dedicated person at the NRF to assist institutions after the setting up of the central archive and the building of the portal for at most a year after initial implementation.

Furthermore, for the first 6 months (approximately), a computer programmer/technician will be necessary to set up the central archive and build the portal. Thereafter, technical support can be provided as needed (possibly 10-20% of one technician's duties).

A librarian may in all likelihood not to be needed at the central site.

Management support is necessary and the project should fall under an appropriate umbrella, which includes similar projects hosted by NRF.

### **5.1.2 Training**

The survey results indicate that the following skills were necessary for staff working on the ETD projects:

- Ability to install and use Digital Library software (DSpace, EPrints, ETD-db, etc.)
- Knowledge of Unix/Linux, XML, Perl, MySQL, HTML, PDF
- Programming
- Understanding of metadata and metadata conversion/importing/exporting
- Ability to analyse and understand statistics
- Understanding of principles of OAI and Open Access
- Ability to organise and catalogue ETDs
- Knowledge of copyright issues
- Clerical skills
- Communications skills

In most cases, staff with these skills worked on the ETD projects, without further training. Some respondents indicated that staff acquired knowledge in new areas by attending training workshops and the annual NDLTD-affiliated ETD symposium.

Mailing lists and periodic meetings of staff have aided communication and thus improved the skills of staff.

## **5.2 Technical Requirements**

The central NRF ETD project will require a computer server to host the Metadata Archive and the NRF ETD Portal. With sufficient storage space and redundancy, the cost for this can be estimated (based on survey responses, recent purchases at UCT-CS, and a quote listed in APPENDIX A) at R20 000-62 000, ranging from bare minimum to enterprise-grade hardware. Survey respondents confirmed that the cost of additional archives or hosting of ETD collections is negligible after the initial purchase. Also, there is no major startup cost besides the purchase of a server.

The server will require an Internet connection. Based on current data for ETDs, the service provided by TENET should be adequate for the near future.

In terms of software, Open Source Software can be obtained (and is typically used for ETD projects) for the operating system, Web server, Web applications, Digital Library systems and databases. It is recommended that NRF uses the DSpace [DSpace Federation, 2006] software to host ETD collections for other institutions (based on current expertise in South Africa and the thesis-specific tools developed for DSpace).

### **5.3 Workflow and Planning**

The national project must set up clear instructions for partners on how to set up OAI data providers and thereby cooperate and integrate with the national project [Reeves, 2005; Theses Canada, 2005; Pavani, 2004].

## **6 REQUIREMENTS FOR INDIVIDUAL INSTITUTIONS**

Individual institutions with *existing* ETD projects *should not* require additional staff or equipment to support the national project.

Pavani [2004] states that members of a national ETD project need to implement the shared standards and transfer metadata to the national site regularly (preferably via OAI-PMH). This means that each institution will need to ensure that it has an operational OAI interface, conforming to the national project's guidelines.

From a non-technical perspective, individual institutions need to make a commitment to join the national project, agree with its rules and be prepared to commit staff and technical resources to assist with its functioning. Generally, the issues around policy in relation to IP and copyright need to be addressed well in advance, due to the different perceptions of ETDs by institutions

The requirements for institutions to prepare and submit theses and dissertations electronically at local institutional level are well understood and documented, and these have been implemented at a third of South African institutions. As such, the need for the elaboration of the latter as part of the current study seemed superfluous, as it is the sort of information which could easily be exchanged at subsequent stakeholder meetings. Detailed requirements for the establishment of individual ETD collections can also be found in the Guide to Electronic Theses and Dissertations<sup>10</sup>, a UNESCO-sponsored effort to document the process of setting up and managing ETD collections.

Nevertheless, while collecting data on national project technical requirements, respondents were also asked about institutional projects. Based on their responses, it is apparent that individual institutions wanting to set up ETD projects will require staff and equipment similar to that discussed in Sections 5.1 and 5.2, with the following major differences:

- ❖ A librarian needs to be part of the process to catalogue and archive ETDs locally.

---

<sup>10</sup> <http://www.etdguide.org/>

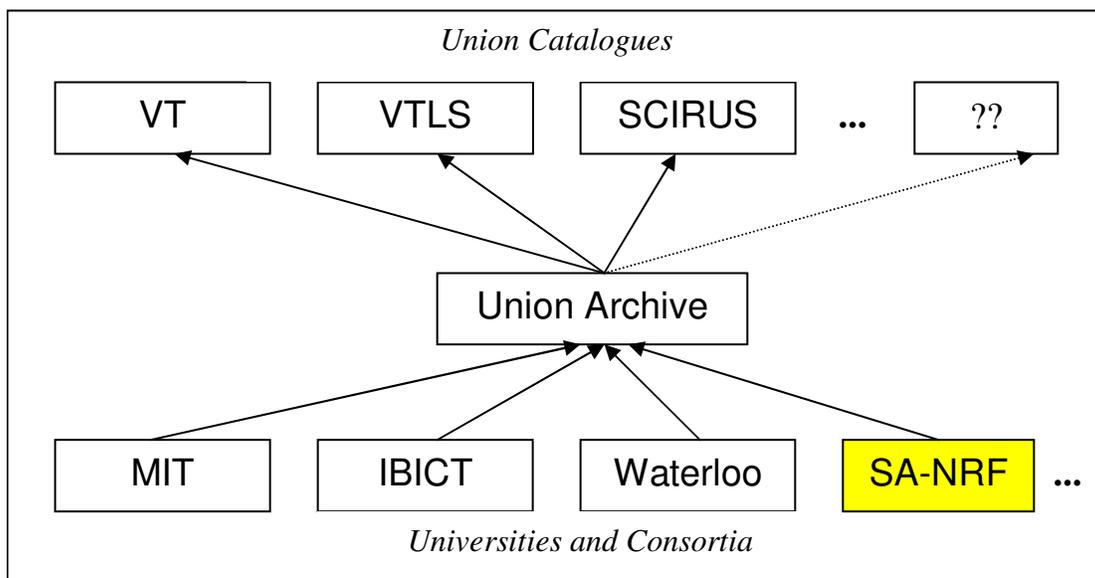
- ❖ Administrative support is necessary to collect documents and process paperwork and payments (if, for example, students pay an archival fee).

As an aside to this report, it should be noted that this report has not set out to elaborate on the status quo of current ETD implementations in South Africa, nor has it set out to establish the reasons behind the lack of implementation of ETD projects at institutions. Both of the aforementioned are seen as being beyond the scope of the study's current remit.

## 7 INTEROPERABILITY WITH NDLTD UNION CATALOG

At an international level, the national project should cooperate with the NDLTD's Union Catalogue project, by contributing South African metadata into this global discovery system.

The NDLTD Union Catalog Project comprises two components: a single centrally managed Union Archive and a collection of independent and distributed Union Catalogues [Suleman and Fox, 2002]. Figure 5 illustrates the relationship among these entities and the individual institutions and national/regional projects.



**Figure 5. Union Archive, Union Catalogues and University ETD Collections**

The NDLTD Union Archive is a metadata repository that periodically collects metadata from various partner sites around the world with the intention of republishing this metadata to service providers such as search engines. The Union Archive is currently managed and hosted by OCLC in Dublin, Ohio, USA, as a free service to NDLTD members. The architecture of the Union Archive is such that it can be relocated or replicated easily, thus stability of the central site is not an issue. Also, OCLC as an organisation provides critical services to the international library community so are in a prime position to manage the Union Archive over a sustained period of time.

The Union Catalogues are service providers who provide discovery and access services based on the data from the Union Archive. Currently, Union Catalogues are being

maintained by Virginia Tech, VTLS and SCIRUS (Science Index service run by Elsevier).

After collecting metadata from each of its member institutions, the South African National ETD project can republish this metadata through an OAI-PMH data provider hosted by the national project. This will allow any external party to get a copy of the metadata only. By registering this data provider with NDLTD (as per instructions published by NDLTD [Suleman, 2004]), metadata from South African institutions will be included in the NDLTD Union Archive on an ongoing basis, and will thus be discoverable by the international community of researchers using any of the Union Catalogues.

In addition, NDLTD is currently negotiating with Google Scholar to include ETD metadata and full-text in its indices. By participating in NDLTD and its Union Archive, these benefits will be automatic for the South African project.

## **8 SUSTAINABILITY AND PRESERVATION**

Sustainability and preservation entail fiscal, policy and technological considerations. Not many current national initiatives have preservation as their primary consideration [McMillan, et al., 2005]. The NRF will need to consider the perpetuation of its role as project funder, noting that elsewhere in this report it was found that survey respondents considered that funding beyond the initial capital outlay was at a minimum. Nevertheless, the host of the national project has to provide for at least a minimal level of technical and administrative support. Preservation is also recognised as an unknown for electronic documents, where paper counterparts are known to last hundreds to thousands of years. Members of NDLTD have been actively researching solutions to this problem and the South African National ETD project can and should interface with such projects. The role of the National Library of South Africa (NLSA) should also be considered here, as one of the key functions of any national library is preservation – ETD projects in some countries are strongly connected to their national libraries (e.g., Germany, Portugal). Furthermore, CHELSA rightly sees the NLSA as a key stakeholder and partner in the national ETD project.

In terms of policy, ETD projects are most effective when there is a compulsory submission requirement at each institution. In the case of Australasian Digital Theses, optional submission yielded 12% of documents in electronic format while mandatory submission naturally yielded 100% [Wells, 2005]. In forming policy around ETDs, IRs and Open Access, NRF needs to consider the importance of “filling the archives”. Moreover, it is important to bear in mind the recommendations of the ICSU/CODATA workshop mentioned under Section 3.1 above, particularly recommendations 1.2, 3 and 3.3. which refer to mandated long-term curation of research outputs, cost-effective preservation of publications and the archiving of national heritage items, respectively. These are important policy recommendations vis-à-vis broad definitions of the sustainability and preservation of scholarly research.

A digital preservation policy is usually also a requirement for a national archive. This policy will need to evolve in the current landscape of ongoing research into preservation mechanisms.

In terms of the technology, current research efforts will need to be tracked, in collaboration with international partners, in terms of format migration, emulation and standardisation for ETDs.

## 9 RECOMMENDATIONS

This report has attempted to discuss the requirements and feasibility of a national ETD project in South Africa. It is abundantly clear from literature, feedback from individuals and the documented examples of projects in other countries that a national ETD project is both feasible and highly desirable for the benefits outlined in this document.

As a concrete way forward, it is recommended that

- NRF (Executive Director KM&S) and CHELSA should establish a Reference Group/Steering committee with the general responsibility of compiling the Terms of Reference for the National ETD project and the way forward, and a Technical Committee (tasked with establishing standards, best practices and a portal).
- NRF should establish a national ETD portal with discovery and access services, and possibly archival services.
- NRF should in principle, be amenable to hosting ETD collections for institutions that are not capable of doing this for themselves.
- NRF should host a meeting of representatives of South African universities to discuss and clarify the ownership and rollout of a National ETD Project and determine which institutions need direct hosting support.
- The national ETD portal should be so setup that it can be easily moved to another location should the need ever arise.
- Technical aspects of a national ETD project could be addressed in collaboration with or as a project initially commissioned to a research laboratory with relevant expertise, such as the Advanced Information Management laboratory at the UCT Department of Computer Science – headed by one of the co-authors of this report. This could accelerate the process and ensure that current best practices in the creation of Digital Library systems are operationalised, while also ensuring a strong link between production services and related research and supporting the training of students wherever possible.
- NRF should budget for equipment and staff support for the first year and subsequent years of the project. In-house IT staff may be appropriate for ongoing support after year 1, but initial work may need to be outsourced on a consultative basis.
- NRF might need to consider how its legacy NEXUS bibliographic database of theses and dissertations will be migrated to the proposed national ETD project; and further, how to go about obtaining the associated fulltext if this is deemed important.
- NRF should bear in mind that though the remit of this study has been ETDs, the possibility strongly exists that institutions may want to implement institutional

repositories (IRs) hosting the research output of academics and researchers, in addition to graduate research output; and that the NRF should anticipate its role in this regard

- NRF should join the NDLTD and actively participate in the international movement to support ETD collections.

It is hoped that these recommendations will form the basis for discussion, and a deliberate and concerted set of activities, towards the establishment of an ETD project, as a matter of some urgency.

## 10 ACKNOWLEDGEMENTS

We wish to thank NRF for their financial support and CHELSA for commissioning this long-overdue study; Andrew Kaniki, Refiloe Mabaso, Ellen Tise, Felix Ubogu, Henda van der Berg and Susan Veldsman, and their respective colleagues, for commenting on draft versions of this document; and the various NDLTD members who responded to the survey on ETD project requirements.

## 11 REFERENCES

Academy of Science of South Africa, 2006. Report on a strategic approach to research publishing in South Africa (March 2006).

Atkins, Anthony, Edward Fox, Robert France and Hussein Suleman (2006) ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations -- version 1.00, revision 2, NDLTD. Available <http://www.ndltd.org/standards/metadata/current.html>

Campbell, Debbie (2005) Easy to do, a brief history of federated harvesting in Australia, ETD 2005: 8<sup>th</sup> International Symposium on Electronic Theses and Dissertations, Sydney, Australia, 28-30 September 2005. Available [http://adt.caul.edu.au/etd2005/pres/GalleryOne03Fri/04\\_A2\\_0830-1030\\_Debbie\\_Campbell.ppt](http://adt.caul.edu.au/etd2005/pres/GalleryOne03Fri/04_A2_0830-1030_Debbie_Campbell.ppt)

Committee on Data for Science and Technology (CODATA) (2006). Executive Summary and Recommendations: Strategies for Permanent Access to Scientific Information in Southern Africa: Focus on Health and Environmental Information for Sustainable Development. An International Workshop held in Pretoria, South Africa, 5-7 September 2005 Available <http://stardata.nrf.ac.za/html/ExecSummaryFinal.doc>

Dublin Core Metadata Initiative (2004) Dublin Core Metadata Element Set, Version 1.1: Reference Description, DCMI Recommendation. Available <http://www.dublincore.org/documents/dces/>

International Council for Science (ICSU) (2004). Scientific Data and Information : report of the CSPR Assessment Panel Available [http://www.icsu.org/Gestion/img/ICSU\\_DOC\\_DOWNLOAD/551\\_DD\\_FILE\\_PAA\\_Data\\_and\\_Information.pdf](http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf)

- McMillan, Gail, Catherine M. Jannik and Robert McDonald (2005) A practical, working and replicable approach to ETD preservation, 8th International Symposium on Electronic Theses and Dissertations, Sydney, Australia, 28-30 September 2005. Available <http://adt.caul.edu.au/etd2005/papers/092McMillan.pdf>
- Kaniki, Andrew, and Henda van den Berg (2003) Electronic information resources and the move towards the provision of full text through and by the NRF to enhance knowledge production, Workshop on the role of electronic information resources in enhancing and supporting research output in higher education institutions (HEIs), National Research Foundation, Pretoria, 30 September 2003. Available <http://stardata.nrf.ac.za/doc/etdWSKanikipaper.doc>
- Lagoze, Carl, and Herbert Van de Sompel (2001) The Open Archives Initiative: Building a low barrier interoperability framework, in Proceedings of First ACM/IEEE Joint Conference on Digital Libraries (Roanoke, USA, 2001), ACM Press, New York, NY, 2001, 54-62.
- Lagoze, Carl, Herbert Van de Sompel, Michael Nelson and Simeon Warner (2002) The Open Archives Initiative Protocol for Metadata Harvesting, Open Archive Initiative. Available <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- Leiner, B. (1998) The NCSTRL Approach to Open Architecture for the Confederated Digital Library, D-Lib Magazine, Vol. 4, No. 11, December 1998. Available <http://www.dlib.org/dlib/december98/leiner/12leiner.html>
- McMillan, Gail (2004) Implementing ETD Services in the Library, chapter in Fox, E. A., S. Feizabadi, J. M. Moxley, and C. R. Weisser (eds), Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians, Chapter 22, pp. 319-329, Marcel Dekker, Inc..
- Müller, Eva, Uwe Klosa, Stefan Andersson and Peter Hansson (2003) The DiVA Project - Development of an Electronic Publishing System, DLib Magazine, Vol. 9, No. 11, November 2003. Available <http://www.dlib.org/dlib/november03/muller/11muller.html>
- NISO (2003) Information Retrieval (Z39.50): Application Service Definition Protocol Specification, Z39.50-2003, ANSI/NISO. Available <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf>
- Pavani, Ana (2004) Consortia & Networking to Build ETD Programs, ETD 2004: 7<sup>th</sup> International Symposium on Electronic Theses and Dissertations, Lexington, Kentucky, USA, 3-5 June 2004. Available <http://www.uky.edu/ETD/ETD2004/pavani/rookie.ppt>
- Pavani, Ana, and Sueli Maffia (2005) BDTD – The Brazilian National ETD Project, ETD 2005: 8<sup>th</sup> International Symposium on Electronic Theses and Dissertations, Sydney, Australia, 28-30 September 2005. Available <http://adt.caul.edu.au/etd2005/papers/114Pavani.pdf>
- Reeves, Sharon (2005) The development of open access to ETDs in Canada: a partnership between Canadian Universities and Library and Archives Canada, ETD 2005: 8<sup>th</sup> International Symposium on Electronic Theses and Dissertations, Sydney,

- Australia, 28-30 September 2005. Available <http://adt.caul.edu.au/etd2005/papers/035Reeves.pdf>
- Suleman, H. (2004) NDLTD Union Catalog Project, chapter in Fox, E. A., S. Feizabadi, J. M. Moxley, and C. R. Weisser (eds), *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, Chapter 6, pp. 73-77, Marcel Dekker, Inc..
- Suleman, Hussein, Anthony Atkins, Marcos A. Gonçalves, Robert K. France, Edward A. Fox, Vinod Chachra, Murray Crowder and Jeff Young (2001) *Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 1: Mission and Progress and Part 2: Services and Research*, *DLib Magazine*, Vol. 7, No. 9, September 2001. Available <http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html> and <http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html>
- Suleman, Hussein, and Edward A. Fox (2002) *Towards Universal Accessibility of ETDs: Building the NDLTD Union Archive*, at Fifth International Symposium on Electronic Theses and Dissertations (ETD2002), Provo, Utah, USA, 30 May-1 June 2002. Available [http://www.husseinsspace.com/publications/etd\\_2002\\_paper\\_union.pdf](http://www.husseinsspace.com/publications/etd_2002_paper_union.pdf)
- Theses Canada (2005) *How to Set Up an ETD Submission Program*. Website <http://www.collectionscanada.ca/thesescanada/s4-210.1-e.html>
- Tise, Ellen (2005) *ETD development In Southern Africa: challenges and innovations*, ETD 2005: 8<sup>th</sup> International Symposium on Electronic Theses and Dissertations, Sydney, Australia, 28-30 September 2005. Available <http://adt.caul.edu.au/etd2005/papers/116Tise.pdf>
- Ubogu, Felix (2004) *ETD Pilot Project Final Report*, University of the Witwatersrand, March 2004. Available <http://www.witsetd.wits.ac.za/ETD-db/pilotproject/finalreport.doc>
- University of Southampton (2005) *EPrints Free Software*. Website <http://www.eprints.org/software/>
- Van de Sompel, Herbert, Carl Lagoze, Michael Nelson, Simeon Warner, Patrick Hochstenbach and Henry Jerez (2004) *Specification for an OAI Static Repository and an OAI Static Repository Gateway*, Open Archive Initiative. Available <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>
- Van der Walt, Wynand (2006) *Personal interview*. University of the Western Cape, South Africa, 21 April 2006.
- Wells, Andrew (2005) *The Australian perspective: innovation and stewardship in building a key national research repository - the Australian Digital Theses Program*, presented at ETD 2005: 8<sup>th</sup> International Symposium on Electronic Theses and Dissertations, Sydney, Australia, 28-30 September 2005. Available [http://adt.caul.edu.au/etd2005/pres/Ritchie01Wed/05\\_C1\\_1315-1455\\_Andrew\\_Wells.ppt](http://adt.caul.edu.au/etd2005/pres/Ritchie01Wed/05_C1_1315-1455_Andrew_Wells.ppt)

## APPENDIX A

The following is a quotation (from Datacentrix, 25 May 2006) for a server suitable as a national ETD project union catalogue and portal, given the requirements such as availability, redundancy, and user base.

### IBM X346 and 146GB Drives for EXP400

The X346 with 6x300GB drives would give you approximately 900GB usable on RAID 1 and 0. If you need more you would need to go external.

The 146GB drives have a different part number as this is part of IBM's ROHS (removal of harmful substances) initiative – this is the non-harmful direct replacement.

IBM x346 Server				
Qty	Model	Description	Unit Price	Ext. Price
1	8840EFG	Express x346, Xeon 3.2GHz/800MHz, 2MB L2, 2x512MB, SR 7k, 8x-24x, 2x625W p/s,Rack (73GB Drives Removed)	R 19,040	R 19,040
1	40K2505	3.2GHz Processor	R 7,552	R 7,552
1	39M5821	1 GB (2x512MB KIT) PC2-3200 CL3 ECC DDR2 SDRAM RDIMM	R 2,772	R 2,772
6	40K1025	300GB Hot-Swap 3.5" 10K RPM Ultra320 SCSI HDD	R 5,377	R 32,260
				<b>R 61,624</b>

146GB Drives				
Qty	Model	Description	Unit Price	Ext. Price
3	40k1024	146GB Hot-Swap 3.5" 10K RPM Ultra320 SCSI HDD	R 3,234	R 9,703
				<b>R 9,703</b>

**Notes:** Pricing quoted above excludes VAT, is valid for 7 days.  
Pricing is quoted at a rate of exchange of \$1 = R 6.55  
Pricing is subject to exchange rate fluctuations.

## APPENDIX B

The following are unprocessed responses to the indicated non-demographic questions on a brief survey on ETD collections at institutional and national levels.

*What services are provided for your local ETD collection (using your institution's equipment and staff)?*

Total responses to question: 11

Submission of ETDs by local students/administrative staff	10
Review of ETDs by examiners/administrative staff	7
Storing of metadata for local ETD collection	11
Storing of fulltext for local ETD collection	11
Hosting of search and retrieval services for local ETDs	11
Backup/Preservation/Management of local ETD collection	10

*How many students submit theses/dissertations each year (approximately)?*

700
450
40
100-150
170 (approx 2/3 submit ETDs)
100
600
450
2005: 143 theses; 315 dissertations
800

*How many staff are employed to work on the ETD project? Use fractions or percentages to indicate the amount of staff time dedicated to the ETD project. (e.g., 1 computer technician, 1/2 librarian and 1/2 manager)*

1/2 computer technician, 1/2 librarian, 1/2 manager
1/2 librarian plus a team of 6 librarians that meets once per month, currently 1/2

computer technician (though his work on the project will be dropping off after the next couple of months), plus advisors in the graduate school who are in charge of theses and dissertations
Total 2 FTE = 2 student assistants (1 FTE), 1/2 library paraprofessional, 1/2 library
1 half-time Librarian
NONE! This responsibility was added on to my duties as Library ICT Coordinator. I can spend at most 1/3 of my time on Repository management/administration, including ETDs, a most unsatisfactory situation. * 5% computer technician. (Library ICT Expert provides some support) * Systems staff at local IT Division provide technical support (e.g. software installations, upgrades, configuration, more complex programming changes) * Library management and Information Services Librarians are involved with policies and advocacy. * No assistance with metadata entry and no administrative assistance. * Ideally 1 full-time manager/administrator with an ICT background and 1/2 librarian
1 computer technician, 1 librarian, 1 clerk
Once the system was up and running, the move to electronic has resulted in an overall decrease in staff time devoted to theses and dissertations. As primary administrator and librarian supervisor of the digital content management system, metadata enhancement, etc. I have been the primary staff devoted to ETDs. I would approximate the time I've devoted to the ETD project as .1 FTE.
1/2 computer technician, 1 information technology consultant, 1/4 librarian, 1 manager.
1 administrative assistant who does a lot of the management as well 1 hour per week (maximum) of 1 computer technician 1/40 2 days per month of one senior manager 1/10 We also outsource dissertations that still need to be PDFed
1/3 library systems administration 1/3 library web administration 1/3 library liaison 1/2 Graduate School reviewer

*What specific skills were necessary among the project staff to set up your ETD project? Comment also on any staff training to support your ETD project.*

The ETD project began approximately a year before Texas Tech hired me, but it stalled for various reasons before the system interface was ready for public use (though the students were submitting ETDs). I took over the project a few months ago, and we're in the process of migrating the collection to a new system. As far as I know, the project has entailed the following: Researching issues surrounding ETDs and communicating with the Graduate School as well as the entire University to gain buy-in for student-submitted ETDs (not all faculty are yet convinced that having electronic-only copies is a good idea). Researching systems for accepting and storing ETDs, evaluating them, and choosing one to use. Installing the chosen system in a Unix-based environment. Working with the Graduate School's theses/dissertations advisors to set up policies and procedures for ETDs. Currently,
---

as project lead, I have implemented steps to clean up the ETD metadata (which is student-submitted), and this has entailed some amount of programming on our current system, and I've worked with one or two of our IT staff to get a new system set up. Now I am working on programming something that will allow ETDs to be submitted/processed in the original system, and then exported to the new system to be made publicly accessible. I'm still working with the Graduate School to work out a few kinks in the process and to allow the Music School to submit ETDs that have sound and video files as components. Again, as far as I know, specific training for the ETD project was never required, except that one of our associate deans attended an ETD conference a couple of years ago, and I may be attending the ETD 2006 conference in June.

Training for faculty included attending two ETD conferences, ALA discussion groups, and lots of reading. Training for library paraprofessional in using the equipment, and analyzing statistics. Training for students in using equipment and recording statistics.

\* Knowledge of copyright \* Knowledge of ftp, file retrieval and management \* HTML skills \* High level cataloguing skills

\* Repository systems installation, configuration and maintenance requires thorough knowledge of Linux (we use FreeBSD) and XML, PERL and MySQL programming skills. \* The repository administrator need some understanding of XML and XHTML, and must be familiar with the relevant metadata schemes. \* Experienced with Adobe Acrobat software - to assist candidates with PDF conversion; \* Training organised by SASLI was very useful except we don't use dSpace. A good understanding of the principles & history of open access, the OAI and Intellectual Property Rights also important. \* Training in exporting and importing of metadata would be useful skills.

Computer skills are necessary to maintain the ETD-db application. Librarian skills are required to effectively organize ETD information. Clerk skills are necessary to keep day-to-day operations flowing smoothly.

Training on DSpace software.

1) Our initial home-grown system in 1998 was implemented on Microsoft Access database platform. Cold Fusion scripting was used for automated interfaces including file uploads to Proquest, submission and review email correspondence and search queries. The former director of our academic computing unit set this up on his own. 2) Our current home-grown system was migrated from the above (MS Access db) to an Oracle based platform. Our Office of Information Technology was involved in this process. The initial migration project required part time assistance from 1 information systems head, 3 Oracle system developers, 1 Java developer, 1 Web design consultant and 1 library ETD Program Coordinator/Manager. Technical support for students (to create and submit ETDs) is also provided by our Office of Information Technology. One campus-wide workshop is held each semester for students, faculty and staff. Several hands-on workshops in the lab are held each semester for students. During the last 2 weeks of each semester a walk-in clinic is held for students to complete pdf, etc. file production and online submission.

Occasional library staff training sessions are held, particularly for our reference and interlibrary loan staff.
Skills for - system installation and maintenance - web site creation and maintenance - secretarial work including document formatting, dealing with other media etc. - process/workflow organization - training of students - developing tutorials and other forms of self help
The library staff used the skills they acquired through other library initiatives, e.g., eReserves, online news posting, ejournal hosting, etc.

*How much have you spent on equipment for local ETD projects? If the costs are negligible as part of other projects, answer 'negligible'. Please indicate the currency, e.g., 2000USD, 15000ZAR.*

During evaluation/planning	When the project became active	Each year after that, on average
negligible	negligible	negligible
\$3000usd	\$10,500usd	negligible so far (1st year post-active)
	A\$2000 for computer, A\$500 for scanner, and approx. A\$2000 for a server	approx A\$1125
negligible	15000ZAR (server)	negligible
negligible	negligible	negligible
negligible	negligible	negligible
0USD	30,000USD	0USD
negligible	43140ZAR	
negligible	negligible	negligible

*What services are provided by the(is) regional/national initiative (irrespective of who has provided the funding)?*

Total responses to question: 7

Harvesting of metadata/fulltext from member sites	6
Union collection of metadata	6
Union collection of fulltext	4
Hosting of search and retrieval services for union collections	7

Backup/Preservation/Management of union collections	4
Participation in International projects e.g., NDLTD	5
Standards development	5
Training of staff at member sites	3
Hosting of proxied collections for under-resourced member sites	1

*How many ETDs (metadata and/or fulltext) are processed per year by this regional/national system (approximately)?*

2200
n/a
5000?
1600-1700
I don't know. We are not yet a member of NDLTD and not terribly familiar with what they do.

*How many staff are employed to work on the regional/national ETD project? Use fractions or percentages to indicate the amount of staff time dedicated to the ETD project. (e.g., 1 computer technician, 1/2 librarian and 1/2 manager)*

0.1 librarian 0.1 web developer (This counts ongoing maintenance, not the original development effort.0
1/10 librarian, 1/4 computer technician
At my site, there is just me and I use two part-time Graduate Assistant. I do not know how many work at OhioLINK ETD site. We process about 250 per year.
Unknown
Don't know. We use our own in-house (home-grown) system to collect and distribute ETDs. Some services such as harvesting are available through NDLTD and OCLC.
1 student computer technician 1/4 computer technician 1 information scientist 1/2 librarian 1/3 manager

*What specific skills were necessary among the project staff to set up your regional/national ETD project? Comment also on any staff training to support your ETD project.*

The initiative is very new, so the bugs are still being worked out. ETDs will be just one in a vast number of collections that the TDL will support. We're still working on allowing the TDL to harvest our ETDs. This is part of the reason that we are switching
--

to a new public ETD system--it will allow the TDL to more easily harvest our metadata. I am on the TDL's Metadata Committee, and the head of the library's IT department is on the TDL's Steering Committee. The Metadata Committee developed a MODS-based ETD standard for the TDL, and we're beginning to think about moving beyond the ETD collection, though we're still working on issues such as collective rights management for ETDs and supporting ETDs that have multiple files associated with them.
No know
Unknown
technical skills legal skills library knowledge negotiation skills coordination skills

*What forms of training have been organised by the regional/national project for member sites?*

A couple of times per year the TDL has a meeting for all institutions and individuals involved in the initiative to summarize what's going on and to demonstrate some of the technology that's being developed to support the TDL. Also, because the initiative is going to require that each member institution use a particular piece of software, the institutions that are doing the most development work on that software have extended an open invitation to all members' IT staffs to spend time learning about the software and the modifications that are being made to support TDL collections. Since the TDL initiative is new, training opportunities and forums to educate member institutions (as well as the general public) are still very much under development.
listserv and they are working on setting up annual meetings
Unknown
NDLTD - all of the above
annual workshops mailing discussions support

*How much has been spent on equipment for the regional/national ETD project? If the costs are negligible as part of other projects, answer 'negligible'. Please indicate the currency, e.g., 2000USD, 15000ZAR.*

	When the project became active	Each year after that, on average
During evaluation/planning		
negligible	negligible	negligible
Unknown		
negligible		

*General comments, lessons learnt, major stumbling blocks, etc.*

<p>Success of a regional ETD union collection relies on clear workflow and consistent practices at the contributing programs.</p>
<p>Our ETD initiative has been more complicated than it perhaps needed to be--the challenges involved are both technical (ensuring preservation as well as a workflow that will allow students to control the amount of access that their ETDs receive) and political. The entire campus has a stake in the university's theses and dissertations, and trying to implement mandatory ETDs requires an enormous amount of buy-in that can be difficult to get. Fortunately, a lot of this was already done before I began working here.</p>
<p>I am new to my position. I have been here about three months.</p>
<p>ETD projects require cooperation among the groups involved. I recommend reports/meetings every 3-6 months for communication.</p>
<p>*Copyright issues *National site is not considered a secure site, and may not be sustainable long-term. *Initial preparation of theses is time-consuming for library staff. Self-submission is not encouraged to overcome copyright issues.</p>
<p>ETDs have staffing implications and the extra workload should not be under-estimated. Provision should also be made for quality control. These are major stumbling blocks in our case. Communicating information about ETDs to candidates and supervisors are also tricky; they are not always aware of their options and the implications of submitting a electronic copy.</p>
<p>Best of luck. Every aspect of our ETD project has been fulfilling: making student research widely available through search engines via OAI-PMH; enabling permanent access online to students' research; promoting OSU student research to the world; working with motivated faculty to make this happen; involving cataloging and other library technical services staff in devising new workflow; increasing efficiency with which theses and dissertations are processed and cataloged.</p>
<p>You can start out small and cheap and still be 'peppy'. Our initial system was hosted on an old server lying around the office so equipment cost was negligible, and development was completed in the spare time of the former director of our academic computing unit. Now we have a much more robust system on the Oracle platform. We have now moved beyond ETDs to host our institutional repository which includes ETDs, Honors Theses (undergraduate) and scholarly resources archive (faculty and staff publications). In the coming year we will issue a request for proposals to see commercial service provider support in an effort to maintain international standards. Estimated migration cost to be 100,000USD and annual maintenance of 30,000USD. If you have sufficient information technology expertise on campus, you could customize and use share-ware systems such as the Virginia Tech ETD system, D-space or E-prints.</p>
<p>*For new implementations/projects we always advise people to start small with volunteers, do most of the work yourself while you sort out your procedures and possible problems. *Gradually expand once you are able to demonstrate the sensibility</p>

of the project and only then start lobbying for infrastructure and policy changes. \*It is absolutely necessary to make it mandatory at some stage: all of a sudden all the objections disappear and a vast lot of time is saved. \*Make sure that all role players on campus are enlisted and play their parts: IT, student admin... It is equally important for them that it is a success

We have posted lots of information at <http://scholar.lib.vt.edu/theses/data/> Information about equipment, staff, etc. is here <http://scholar.lib.vt.edu/theses/data/setup.html> and <http://scholar.lib.vt.edu/ETD-db/developer/require.html>