

LIFECYCLE MANAGEMENT OF ELECTRONIC THESES AND DISSERTATIONS

IMLS NLG Proposal

Submitted Tuesday, February 1, 2011

LIFECYCLE MANAGEMENT OF ELECTRONIC THESES AND DISSERTATIONS

ABSTRACT

The University of North Texas Libraries (lead applicant) together with the Networked Digital Library of Theses and Dissertations (NDLTD), the Educopia Institute/MetaArchive Cooperative, and the libraries of Virginia Tech, Rice University, Boston College, Indiana State University, Pennsylvania State University, and the University of Arizona propose to develop and disseminate guidelines, educational materials and an associated workshop, and a set of software tools for life-cycle data management and preservation of Electronic Theses and Dissertations (ETDs). This project will take place over a two year period from October 2011 to September 2013, and falls within the Advancing Digital Resources category of the IMLS NLG program.

Intended Audience

The intended audience for this project includes academic libraries currently managing or prospectively considering programs for ETD preservation. Colleges and Universities have been steadily transitioning from traditional paper/microfilm to digital ETD submission, dissemination, and preservation processes. Increasingly, colleges and universities worldwide are accepting and archiving only electronic versions of their students' theses and dissertations. While this move from print-based to digital-based theses and dissertations greatly enhances the accessibility and sharing of graduate student research, it also raises grave concerns about the potential ephemerality of these digital resources. How will institutions ensure that the electronic theses and dissertations they acquire from students today will be available to future researchers? We need to better understand, document, and address the preservation challenges presented by ETDs to ensure that colleges and universities have the requisite knowledge to properly curate these new collections.

Project Activities and Products

1. *Dissemination of Guidance Documents for Lifecycle Management of ETDs:* These products will provide assistance to information professionals seeking to effectively manage the lifecycle of their ETDs. Based on collaborative research between members of the worldwide NDLTD consortium and the MetaArchive Cooperative, we will develop a series of concise and informative documents concerning the issues of specific concern to the curators of ETD collections, and address the considerations that ETD curators must understand in preservation solutions for data management and preservation of ETD collections.

2. *Production of ETD Lifecycle Management Tools:* The project will develop and disseminate a set of software tools to address targeted needs in managing ETDs throughout their lifecycle. These tools will be created as completely modular micro-services that can be used alone or incorporated into larger repository systems. These micro-service tools will be targeted at individual functions such as ETD format recognition and PREMIS metadata event record-keeping.

3. *Creation of Educational Materials and Associated Workshop:* A set of educational materials on the topic of lifecycle management of ETDs will be prepared and made freely available. These materials will be utilized in a workshop that will be offered in the second year of the project. Materials will include curriculum syllabi, training handouts, PowerPoint presentations, exercises, and other relevant items.

Project Outcomes

The success criteria for this project include the following measurable changes in audience members:

- A. ETD lifecycle management knowledge and skills gained by workshop attendees and others nationwide that use the project products will be assessed.
- B. Utilization of ETD lifecycle management software tools created will be assessed for a nationwide group of institutional users who download and install the tools.
- C. Attitudes and adoption behavior regarding project products of NDLTD members concerning the guidelines will be tracked after dissemination of the guidelines and use cases.

1. ASSESSMENT OF NEED

One of the most important new responsibilities for academic libraries that has emerged in recent years is curatorial responsibility for electronic theses and dissertations. As the initial research products created by new scholars to demonstrate competence during the process of graduate degree certification, theses and dissertations are very important intellectual assets regularly generated by universities. Because virtually all theses and dissertations are now created as digital products with new preservation and access characteristics, a movement toward Electronic Theses and Dissertations (ETD) curation programs in both U.S. institutions and abroad began in the early 1990's and has continued to this day. There are many articles that document this movement (note: all citations in this proposal reference the *Appendix A: Bibliography* in the supplemental documents of this proposal). For example, the Coalition for Networked Information (CNI) recently studied the history of ETDs and graduate education and conducted an international survey concerning ETDs that examined linkages between the growth of ETD programs, institutional repositories, open access and other important trends in higher education. (Lippincott and Lynch, 2010) Additional key issues identified in the CNI survey are questions and uncertainty within institutions concerning embargoes of ETDs, format considerations for ETDs, costs of ETD programs, and the role of libraries in working with graduate schools to maximize benefits of ETD programs for students.

A basic point made by the CNI study and virtually all the literature on the ETD movement reviewed in the bibliography is that colleges and universities have been steadily transitioning from traditional paper/microfilm to digital submission, dissemination, and preservation processes. Increasingly, academic institutions worldwide are now accepting and archiving only electronic versions of their students' theses and dissertations, especially in thesis and dissertation archiving programs operated by academic libraries. While this steady transition in curatorial practice from print to digital theses and dissertations greatly enhances the accessibility and sharing of graduate student research, it also raises grave long-term concerns about the potential ephemerality of these digital resources.

This project proposal focuses on answering the question: *How will institutions address the entire lifecycle of ETDs, ensuring that the electronic theses and dissertations they acquire from students today will be available to future researchers?* This is the context for this proposal, which uses the phrase *lifecycle management of digital data* in the broad sense defined by the Library of Congress to refer to the "progressive technology and workflow requirements needed to ensure long-term sustainability of and accessibility to digital objects and/or metadata" (Library of Congress, 2006), as well as in the more detailed senses of the digital lifecycle management model as articulated by the Digital Curation Centre in the UK. (Higgins, 2008)

In order to unpack this complex issue and assess the needs of libraries that might be addressed through a project focused on lifecycle management for ETDs, leaders of the Networked Digital Library of Theses and Dissertations (NDLTD) and the MetaArchive Cooperative conducted a series of investigations during 2008-2010. (Note: Organizational profiles of the NDLTD and MetaArchive are provided in *Appendix C* of the supplemental documents of this proposal.) These efforts included surveys, an investigative pilot project, and several meetings of the leadership of the two groups, which are each closely concerned with different aspects of preserving ETDs.

In order to assess needs and the current status of the field, a survey was conducted in 2008 by the MetaArchive Cooperative and NDLTD to examine ETD practices and associated concerns in institutions either currently engaged in ETD programs or considering such preservation service programs. The survey received 95 responses, primarily from academic institutions that were providing or strongly considering collection of ETDs and associated ETD services. (McMillan, 2008) The survey asked questions about whether or not the institution accepts ETDs (88% of respondents did accept ETDs), the nature of ETD services, size and formats in the resulting collections, software platforms used, and other questions pertaining to ETD service programs. The most remarkable finding from this survey was *that 72% of responding institutions reported that they had no preservation plan for the ETDs they were collecting.* The survey also collected information about what

information institutions would need to make decisions concerning ETD preservation programs. The responses to this survey led the same researchers to conduct a follow-on survey in 2009 that probed more deeply into digital preservation practices and concerns. (Skinner and McMillan, 2009). This survey included questions concerning institutional policies, knowledge and skills needed for digital preservation activities, level of desire for external guidance and expertise in digital preservation, and perceptions about relative threat levels of different factors in the long-term survivability of digital content. This proposal has especially been informed by the responses in these surveys that identified needs of practitioners in the field.

Based on findings from these surveys, the MetaArchive Cooperative and the NDLTD undertook a joint pilot project in 2008-2010 to further explore and understand issues highlighted in the surveys and to respond to concerns of their respective memberships about preservation of ETDs. In the course of this pilot project, a group of institutions that are members of both organizations (including Virginia Tech, Rice University, Boston College, and others) worked together to discuss, analyze, and undertake experiments in different aspects of lifecycle management of ETDs, and to identify problem areas experienced by multiple institutions. The pilot project group also explored the literature to better understand what has been published to date on different digital lifecycle management topics, and how such publications relate to ETDs. During this pilot project, as another means of assessing needs, Gail McMillan of the NDLTD and Martin Halbert of the MetaArchive Cooperative asked a large number of ETD program leaders about their concerns in the area of ETD lifecycle management during workshops conducted at the three annual ETD conferences hosted by the NDLTD from 2008-2010. Findings from the pilot project analysis and workshop inquiries were reviewed and discussed at three joint planning meetings of the NDLTD board and MetaArchive leadership during this period. This joint planning group decided to proceed with this proposal based on these discussions.

The surveys, pilot project, and leadership discussions all highlighted needs for a cluster of related informational documents and tools to address the broad question of lifecycle management of ETDs. These needs are present in virtually all categories of academic institutions within the United States and internationally. There is need for guidance documents in a variety of specific ETD lifecycle management topics, need for specific software tools to address gaps in services, and need for a range of different types of educational materials to advance the capabilities of institutions that administer ETD service programs. These needs are explicated in the remainder of this proposal in the descriptions of planned products of this project. The planning group strongly feels that as a field we need to better understand, document, and address the challenges presented in managing the entire lifecycle of ETDs in order to ensure that colleges and universities have the requisite knowledge to properly curate these new collections. The intended audience for this project therefore includes the entire range of academic libraries currently managing or prospectively considering programs for ETD preservation and access.

The planning group also believes that other efforts have not addressed the needs identified in this project. While there have been previous efforts to assemble materials to advise institutions concerning ETD programs, the majority of these efforts have been aimed primarily at informing universities of the benefits of ETD programs and motivating campus stakeholders to create such programs, together with general descriptions of what the basic characteristics of ETD programs are and what steps are necessary to successfully propose them. Nor do commercial offerings such as those provided by ProQuest and other vendors resolve the local curatorial needs of institutions that could be met through the proposed products of this project. While vendor services provide important supplements to the capabilities and knowledge of academic librarians, they are not a replacement for informed and prepared academic librarians in this critically important role.

Fundamentally, if academic libraries are to fulfill their responsibilities as long-term custodians of the intellectual outputs of their institutions, they need key guidance documents, software tools, and educational materials. This project seeks to address these needs.

2. NATIONAL IMPACT AND INTENDED RESULTS

The needs assessment process undertaken by the planning group for this project identified a cluster of informational resources needed to advance the acquisition, organization, use, management, presentation, and preservation of ETDs, as well as the need for tools to enhance access, use, and management of ETD digital assets over their entire life cycle. This section will describe these needed resources in more detail as three categories of results for the project. Each of the three categories will also articulate the nationally beneficial impacts of providing these resources.

A. Dissemination of Guidance Documents for Lifecycle Management of ETDs

The needs analysis identified a variety of guidance documents essential to information professionals seeking to effectively manage the lifecycle of their ETDs. This project will produce the following documents, which will address areas of special interest identified by ETD program planners, managers, and stakeholders. These documents will be made freely available through a creative commons license (either the Attribution-NonCommercial-ShareAlike or other license acceptable to IMLS). All of these documents will be carefully researched, and notable experts will be interviewed where appropriate.

1. *Briefing on Access Levels and Embargoes of ETDs*: One of the most contested topics in ETD program planning is the question of ETD embargoes and levels of access restriction, as evidenced by both the NDLTD/MetaArchive surveys and the 2010 CNI survey previously cited. This document will directly address this difficult issue, about which stakeholders are particularly concerned. Thus far, there is no consensus on “good practices” in the field concerning embargoes/restrictions. This briefing will lay out the issues and the different sides of this argument for those considering an ETD program, clearly articulating the pros and cons claimed by proponents for and against embargoes. The point of this document is not to settle the issue, but provide prospective ETD programs with clear and concise information about the ramifications of campus policy decisions for or against different kinds of access restrictions.

2. *Briefing on ETD Copyright Issues and Fair Use*: This document will also provide clarifying information concerning the poorly understood and often hotly contested issues of copyright and fair use in ETDs. There is so much uncertainty and fear over this issue that it has in some cases prevented campuses from implementing ETD programs at all. This briefing will be developed in close reference with the embargo briefing, as the issues are often linked. Both briefings will address publisher concerns and issues squarely. Again, the point of this briefing is to provide a concise summary of the relevant information for stakeholders, with documented citations.

3. *Guidelines for Implementing ETD Programs - Roles & Responsibilities*: This document will provide information to help institutions understand the functions of different stakeholders in an ETD program, thereby streamlining the process of implementing or overhauling ETD programs. This document will provide guidance for understanding roles and responsibilities, together with frank acknowledgements of the concerns that different stakeholders may have, and how ETD programs have addressed these issues as well as the ramifications of *not* specifying roles and responsibilities.

4. *Guidelines for Collecting Usage Metrics & Demonstrations of Value for ETD Programs*: This document will provide guidance for institutions concerning the critical issue of assessment of ETD usage, and how communicating such assessment metrics can demonstrate the benefits of the program to stakeholders. Practical examples of how to document and convey usage metrics will be provided.

5. *Overview of Formats, Complex Content Objects, and Format Migration Scenarios for ETDs*: Additional issues of great concern to many current and prospective ETD programs are how to determine what formats they should accept, and how to manage on an ongoing basis the increasingly complex ETDs that are now being produced by students. This document will seek to concisely set forth the relevant information and

issues for decision makers and practitioners, and to provide scenarios for managing and migrating complex ETDs over their extended lifecycle.

6. *Overview of PREMIS Metadata & Lifecycle Event Record-Keeping for ETDs*: Another issue revealed in the needs assessment process was that most institutions do not have workflows and systems in place to capture the appropriate levels of metadata needed to manage ETDs over their entire lifecycle, often because of a lack of awareness of the issues and ramifications of not maintaining such information. A concise overview of this topic will be prepared to inform stakeholders and decision makers about the critical issues to be aware of in gathering and maintaining preservation metadata for ETDs, not just at the point of ingestion, but subsequently, as ETDs often have transitional events in their lifecycle (embargo releases, redactions, etc.). This overview will be developed in concert with the related metadata tools described below.

7. *Guide to ETD Program Cost Estimation and Planning*: One of the most frequently asked questions identified in the CNI and MetaArchive surveys pertains to the costs and process for setting up an ETD program. This guide will provide institutions with succinct information on costs and planning, laying out the critical paths that many ETD programs have charted, together with advice about strategies. This document will be developed in close concert with the ETD Options Guide described below, and will provide cost-benefit analyses of these scenarios to provide institutions with a range of options to consider for their local needs.

8. *Guide to Options for ETD Programs*: Many institutions are delayed in ETD program planning simply because they do not have a clear understanding of the range of options to consider in implementing an ETD program. Restricted or open access? Implement an ETD repository or lease a commercial service? Who has responsibility for what functions? Like the other documents in this series, this guide will be a descriptive (rather than prescriptive) information resource that seeks to explain the relevant decisions institutions must make and to clarify the pros and cons of different options. This document will reference and integrate all the other documents in the series.

National Impacts: The publication and dissemination of these guidance documents by the NDLTD and other project partners will address key informational gaps that the needs assessment process highlighted, and have a far-reaching impact and results through adoption and adaptation in multiple institutions. Libraries need a core set of concise advisory documents to address practical implementation considerations in ETD programs. The project planners believe that the simple but nevertheless innovative strategy of producing targeted and concise guidance documents on the topics that institutions require will catalytically effect systemic change within and across organizations nationally.

B. Production of ETD Lifecycle Management Tools

The project will develop and disseminate a set of software tools to address specific needs in managing ETDs throughout their lifecycle. These tools will be created as completely modular micro-services, i.e. single function standalone services that that can be used alone or incorporated into larger repository systems. Micro-services for digital curation functions are a new approach to system integration pioneered by the California Digital Library and the Library of Congress (Abrams, 2010), and subsequently adopted by the University of North Texas, Chronopolis, MetaArchive, and other digital preservation repositories. (Minor, 2010)

The micro-services described below will be relatively simple to construct, as they are primarily based on the idea of being able to call other existing open source software tools. The particular functions targeted in these micro-services have been selected to fill specific gaps identified by ETD program leaders in the course of the needs analysis. They are designed as standalone micro-services that can be called via command line or script interfaces in order to ensure that the systems can be easily integrated in existing environments in a modular way. Each micro-service will have clear documentation that will enable implementers to deploy the tool in their own setting. The intent of creating these four micro-services is that they will catalytically enhance existing repository systems being used for ETDs, which often lack simple mechanisms for these functions. By

limiting the scope of these micro-services our hope is to ensure that these services are documented and work well. The micro-service packages produced in the course of this project will include the following tools:

1. *ETD format recognition*: Accurate identification of ETD component format types is an important step in the ingestion process, especially as ETDs become more complex. This micro-service will enable batch identification of ETD files through integration of function calls from the JHOVE2 and DROID format identification toolkits. The micro-service output will be structured in ad hoc tabular formats for importation into repository systems used for ETDs such as DSpace, and the ETD-db software, as well preservation repository software such as iRODS and DAITSS and preservation network software such as LOCKKS.

2. *PREMIS metadata event record-keeping*: One gap highlighted in the needs analysis was the lack of simple PREMIS metadata and event record keeping tools for ETDs. This micro-service will generate PREMIS metadata to track a set of transitions in the lifecycle of particular ETDs. The PREMIS metadata elements that the service will be capable of generating will include all PREMIS Event semantic units, using parameter calls to the micro-service. PREMIS metadata records generated by the system will have some profile conformance options, and documentation will be provided on how to use the metadata in different ETD repository systems.

3. *Virus checking*: Virus checking is an obvious service needed in ETD programs, as students' work is often infected unintentionally with computer viruses. This micro-service will provide the capability to check ETD component files using the ClamAV open source email gateway virus checking software. The micro-service will also be designed such that other anti-virus tools can be called with it. Results of scans will be recordable using the PREMIS metadata event tracking service.

4. *Digital drop box with metadata submission functionality*: This micro-service addresses a frequently sought function to provide a simple capability for users to deposit ETDs into a remote location via a webform that gathers requisite submission information requested by the ETD program. The submission information will generate PREMIS metadata for the ETD files deposited. The micro-service will have the capacity to replicate the deposited content securely upon ingest into additional locations by calling other Unix tools such as rsync, and will also record this replication in the PREMIS metadata.

National Impacts: Although necessarily limited in scope, these tools address key functions that virtually all ETD programs must provide; the potential for adoption of these tools is therefore far-reaching. Project participants will use these tools in their own ETD programs, and will disseminate these tools through the means identified in the project design section. All of these tools will be made freely available (see sustainability section) via a free and open source software (OSS) license, either the Educational Community License Version 2.0 or other OSS license acceptable to IMLS. Finally, although these tools are being implemented specifically for ETD programs, we believe that any of these micro-services could be used in other digital archives contexts. The potential applications of these tools in other settings and scenarios describing how the tools might be adapted for other purposes will be included in the documentation produced for these tools.

C. Creation of Educational Materials and Associated Workshop

A set of educational materials on the theory and practice of lifecycle management of ETDs will be prepared and made freely available through a creative commons license (either the *Attribution-NonCommercial-ShareAlike* or other license acceptable to IMLS). These materials will reference the guidance documents, and will be utilized in a workshop that will be offered in the second year of the project. This workshop will address both theory and practice, with hands-on lifecycle tools training.

Materials will include curriculum syllabi, training handouts, PowerPoint presentations, exercises, and other relevant items. Educational materials produced will include the following items.

1. *Workshop Syllabi*: This set of syllabi will describe the curriculum and learning outcomes of sessions on three areas: a) ETD program planning, b) ETD program implementation, and c) technology deployment. These sessions will each be designed to be taught during either a morning or afternoon footprint of time,

enabling the entire project workshop (and subsequent workshops using these materials) to be taught over the course of a day and a half period. The syllabi will also have information for instructors providing details about how to use the guidance documents, software tools, and other materials in the course of the workshop.

2. *Training Handouts and Exercises:* A body of handouts and exercises will be prepared that can be used by instructors in the project workshop and subsequent workshops. These materials will sequentially present and quiz attendees on information from the guidance documents together with information from other sources in the bibliography, with the goal of preparing workshop attendees to plan and implement (or update) an ETD program and the functions provided by the micro-services developed in this project. These handouts and exercises will be capable of being assembled in a workbook together with the syllabi.

3. *PowerPoint presentations:* A set of coordinated Microsoft PowerPoint slides will be provided to facilitate teaching the above materials.

Workshop: A workshop using these educational materials will be held in the second year of the project, with widespread notice provided in advance of the event targeting institutions that are seeking information concerning ETD programs. Our preference is to hold this workshop in conjunction with another ETD event such as the Texas ETD Association or other ETD meetings in 2013 if possible, but failing this possibility we will host the meeting as a standalone event. Learning outcomes of the workshop will be assessed and reported. Feedback will also be solicited during the workshop about the teaching materials, both from attendees and instructors. This feedback will be used to revise the materials in the final months of the project.

National Impacts: The educational materials and workshop will enable this project to directly reach practitioners seeking information and tools relevant to establishing or updating ETD programs. Further, all educational materials will be promulgated through outreach efforts of the NDLTD and other project partners, thereby achieving a national and international audience of institutions with ETD programs. Mechanisms for more specifically assessing impacts of the educational materials and workshop will be detailed in the next section. We also hope to have a catalytic effect by providing these materials to other groups that may seek to host educational and training events on ETD topics.

3. PROJECT DESIGN AND EVALUATION PLAN

The scope of this project is intended to be a manageable and incremental set of outcomes that will nevertheless have a significant catalytic effect on the field on the national scale, as articulated in the previous section. While the full range of conceivable outcomes of the project cannot be easily measured, we have identified the following completion criteria (what determines that the work was completed) and success criteria (what determines whether or not the project was successful), including specified measurable changes in audience members.

A. ETD Lifecycle Management Knowledge and Skills Increase Targets

Our general aim is to increase ETD lifecycle management knowledge and skills in librarians and other institutional stakeholders. Our project team has at least two means to assess such increase in knowledge and skills: during the assessment of learning outcomes in workshop attendees, and more broadly by following up with individuals that download the guidance documents and other project products. The MetaArchive Cooperative subcontract will include responsibility for these assessment efforts.

Completion Criteria: 1) Learning outcomes for all workshop attendees will be assessed, 2) during the first six months that project documents are made publicly available, individual email addresses will be required to receive documents (after this period the documents will be made available for anonymous download). Follow-up contacts will be made with individuals who receive documents after one month to assess basic questions concerning the knowledge and skills conveyed by the documents. Individual names will not be

recorded during these assessment activities, and only aggregate (rather than individual) scores will be reported to IMLS in project interim report narratives.

Success Criteria: 1) All core ETD lifecycle management learning outcomes set forth in the workshop syllabi must be met by workshop attendees, and anonymized scores must be reported, 2) follow-up assessments with individuals receiving project documents must demonstrate evidence to the satisfaction of the project team and IMLS that on average at least half of all individuals learned the relevant ETD lifecycle management topics covered in the documents they received.

B. Utilization of ETD Lifecycle Management Software Tools

Our general aims regarding the software tools created in this project are that the micro-services be both usable and used. The MetaArchive Cooperative will be responsible for assessing these outcomes as follows.

Completion Criteria: Similar to the strategy described above, during the first six months of availability, receipt of the software tools will require that an individual email address be provided. A one month follow-up with individuals will assess whether or not they found the software tools usable, if they implemented the micro-services, and what feedback they have for the development team.

Success Criteria: Number of sites receiving and utilizing the software tools created will be monitored. Follow-up assessments with individuals must demonstrate evidence to the satisfaction of the project team and IMLS that on average at least half of all individuals found the software usable, and that at least ten sites implemented the software.

C. Attitudes and Adoption Behavior Regarding Project Products

Feedback webforms will be provided on both the NDLTDD and MetaArchive websites. Attitudes, adoption behavior, and other user-reported responses concerning the guidance documents, micro-services, and educational materials will be tracked after dissemination of the guidelines and use cases.

Completion Criteria: This feedback will be summarized and reported to IMLS by the end of the project.

Success Criteria: The feedback from users must demonstrate evidence to the satisfaction of the project team and IMLS that the project was successful in that the community found the information and tools generated by the project useful generally and used the information and tools in catalytic ways. This feedback will also be used to guide the long term maintenance and updating of the documents and software.

4. PROJECT RESOURCES: BUDGET, PERSONNEL, AND MANAGEMENT

The project participants request \$268,457 from IMLS for this project, and will provide \$273,381 in cost match. These figures are detailed in the budget forms and budget justification statements.

This project builds on long-standing successful collaborative relationships between institutional leaders of the MetaArchive Cooperative, the NDLTDD, and the new USETDA (both NDLTDD and USETDA board members have contributed letters of support for this project, see *Appendix E*). Our planning group has a strong sense of the next steps needed to advance ETD program resources nationally, and has proposed this project to move this important work forward. Most of the institutional partners in the project have worked together on many previous occasions. A Project Steering Committee will take responsibility for the research and writing entailed in creating many of the project documents and will be comprised of experts from several of the partner institutions, all of whom are listed below. The Programming Team will take responsibility for software development and will consist of the project programmer, Schultz, and Alemneh, with input from Han, Wang, Henry, Skinner, and Halbert. The following are descriptions of key project personnel and their areas of expertise.

Dr. Martin Halbert (UNT Dean of Libraries and Associate Professor) will act as principal investigator. Halbert will devote 8% of his time as cost-match to the project. He has served as PI for grants and contracts totaling more than \$5M, and has expertise in managing collaborative digital library projects involving both academic institutions and researchers. The UNT Libraries are now planning an overhaul of the long-standing campus ETD program, and the activities of this project will be very synergistic with his responsibilities in leading this effort. He will coordinate the steering committee work on the *Guide to Options for ETD Programs*.

Dr. Katherine Skinner (Educopia Institute Executive Director and MetaArchive Cooperative Manager) will act as co-principal investigator and convener of the project steering committee. Skinner will devote 8% of her time as cost-match to the project. She led the NDLTD/MetaArchive ETD pilot project together with McMillan, and has frequently served as a faculty instructor in digital preservation symposia and workshops. She will coordinate the steering committee work on the educational materials.

Gail McMillan (VA Tech Director, Digital Library and Archives) will serve as a member of the project steering committee. McMillan will devote 5% of her time as cost-match to the project. She is a long-standing member of the NDLTD Board of Directors, as well as a founding member of the MetaArchive steering committee. She led the NDLTD/MetaArchive ETD pilot project together with Skinner. McMillan will coordinate the steering committee work on the *Guide to ETD Program Cost Estimation and Planning*.

Matt Schultz (Educopia/MetaArchive Collaborative Services Librarian) will act as *project manager* for the first 15 months of the project during the period of heaviest activities (document creation, workshop design, etc.) until after the workshop, with 100% of his time devoted to the role of project manager and technical writer as part of the Educopia subcontract. Schultz works on targeted MetaArchive digital preservation contracts on a regular basis, and has served as project manager and technical writer in the past on several occasions. (Skinner and Schultz, 2010) As project manager, Schultz will supervise the programmer in concert with Alemneh.

Geneva Henry (Rice University Executive Director, Center for Digital Scholarship) will serve as a member of the project steering committee. Henry will devote 5% of her time as cost-match to the project. Henry has led many repository and digital library projects in the past. She will coordinate the steering committee work on the *Briefing on Access Levels and Embargoes of ETDs*.

Dr. Bill Donovan (Boston College Digital Preservation Manager & ETD Administrator) will serve as a member of the project steering committee. Donovan will devote 5% of his time as cost-match to the project. He has extensive expertise in visual formats and digital imaging. He will coordinate the steering committee work on the *Overview of Formats* document.

Xiaocan (Lucy) Wang (Indiana State University Digital Repository Librarian) will serve as a member of the project steering committee. Wang will devote 5% of her time as cost-match to the project. She has fresh experience in implementation of ETD programs at ISU, and will coordinate the steering committee work on the *Guidelines for Implementing ETD Programs - Roles & Responsibilities* document.

Dr. Patricia Hswe (Penn State University Digital Collections Curator) will serve as a member of the project steering committee. Hswe will devote 5% of her time as cost-match to the project. She has experience in scholarly aspects of repositories, and will coordinate the steering committee work on the *Briefing on ETD Copyright Issues* document.

Yan Han (University of Arizona Associate Librarian for Research Services) will serve as a member of the project steering committee. Han will devote 5% of his time as cost-match to the project. Han has extensive experience with a variety of digital library standards, and will coordinate the steering committee work on the *Guidelines for Collecting Usage Metrics* document.

Daniel Alemneh (UNT Metadata Librarian) will serve as a member of the project steering committee. Alemneh will devote 5% of his time as cost-match to the project. Alemneh has expertise in metadata and

metrics analysis, and will coordinate the steering committee work on the *Overview of PREMIS Metadata* document. Alemneh will also assist Schultz as on-site manager for the project programmer.

Shannon Stark (UNT Strategic Projects Librarian) will coordinate many project activities during the first 9 months of the project, devoting 10% of her time as cost match to the project. Stark will assist Halbert and Skinner in project startup, communication activities, and convening the first project meeting. She will assist Schultz and others in drafting first versions of project documents, and interviews with topical experts.

The *Project Programmer* will be an experienced digital library technologist, familiar with the concepts of micro-services in the context of lifecycle management of digital assets. Several qualified potential programmers are currently available in the UNT digital projects department, and one will be selected based on availability for a 50% time commitment over the course of the project.

UNT Library Administrative Assistants will provide clerical support for the project with a 5% FTE commitment of a senior administrative assistant as cost match, collating mailings and drafts of documents, purchase and reimbursement of travel expenses for meetings, and transcriptions of notes from conference calls and meetings. Sandra Atchison is one such assistant, but others may substitute as needed.

Work Plan

This project will take place over a two year period from October 1, 2011 to September 30, 2013. *Additional details of the work plan are provided in the Schedule of Completion GANTT chart and Appendix B: Detailed Work Plan in the Supporting Documentation.* The following summary identifies the primary segments of the project, key events, and main associated deliverables.

Preparation/Startup Phase (October 2011 – December 2011)

Activities: This startup phase will include initial preparatory activities. The steering committee will be convened, and a regular conference call schedule will be planned. The project manager and project programmer will be hired. Planning for the first steering committee meeting in the following February will begin, including contacting relevant ETD conferences for coordination on venue. The steering committee will review all project plans and goals, and set out initial work assignments to begin in the following phase. The steering committee and project manager will document the functional requirements for the micro-services. *Deliverables:* Micro-services functional requirements.

Development/Deployment Phase (December 2011 – December 2012)

Activities: This project phase is focused on the development of initial versions of all project deliverables, including documents and software. The steering committee will begin to hold weekly conference calls, bring bibliographies up to date and complete other research updating for all guidance documents. The programming team will create the micro-services iteratively using the scrum development methodology. (Deemer et al., 2009) Outlines and early drafts of all project documents as well as initial micro-services demonstrations will be prepared, reviewed, and discussed at the first full steering committee meeting in February 2012, together with public presentations at the TxETDA or other ETD conference about the project to share information and solicit feedback from the community. Based on feedback received from the assembled steering committee and the community, complete initial drafts of project documents and software will be completed for review at the second steering committee meeting in August 2012. This meeting, held at approximately the half-way mark of the project, will be a key milestone in which all project deliverables will recognizably take form. Planning will also take place at this meeting for the project workshop, to be held the following February in 2013. Based on decisions made at this meeting, the first public-release form of the project documents and micro-services will be developed by December 2012. *Deliverables:* Complete versions of all project documents and micro-services. These versions will be usable in the forms delivered by the end of 2012, but will be revised taking user feedback into account during the next project phase.

Dissemination/Assessment Phase (October 2012 – July 2013)

Activities: This project phase will focus on disseminating and refining the documents and software produced by the project, and assessing the project outcomes. The project deliverables will have a soft release circa December 2012, and be presented at venues such as CNI meetings and other professional conferences (opportunities are noted in the detailed project work plan in the appendices). The soft release of project deliverables will be done with the proviso that individuals requesting the documents agree to provide assessment feedback to the project. The project workshop will be publicized in the Fall of 2012, and held in February 2013 together with the concluding meeting of the steering committee. Members of the steering committee will participate in the workshop as instructors. Assessment activities will take place throughout this phase, and will guide refinements made to the project deliverables in the final six months of the project to respond to community requests for enhancements or changes. *Deliverables:* Semifinal versions of all project deliverables will be produced during this phase. Assessment reports will be collated and prepared.

Reporting/Completion Phase (July 2013 – September 2013)

Activities: This phase will focus on finalizing all deliverables, writing up final project reports, and transitioning the project deliverables to maintenance organizations. *Deliverables:* Final project deliverable updates, and final project reporting.

5. COMMUNICATION PLAN

This project seeks to catalytically advance and foster the national understanding of lifecycle management of ETDs through the dissemination and use of guidance documents, micro-service software packages, and educational materials. This will be accomplished through a variety of means.

Project Websites: The project will maintain two public websites with information about the project and the documents and micro-services produced. The first website will be hosted by the MetaArchive Cooperative, and will be managed by Schultz and Skinner. This website will also serve as the website for the workshop when it occurs, and provide links to presentations and documents used, as well as other associated resources. The second website will be maintained by the ND LTD and managed by McMillan, and will link to all the documents and micro-services produced as a way of broadening the awareness of project deliverables.

Project Workshop: The project workshop that will be held circa February 2013 is one of the key deliverables of the project, and will be a means of directly disseminating project documents and software.

Published Papers and Presentations: Members of the steering committee will publish papers in professional journals and make presentations at conferences on the documents and software produced in this project. These forms of dissemination will be leveraged to make contact with prospective users.

6. SUSTAINABILITY

The results of this project will continue to be supported and cultivated beyond the end of the project by the project partners, especially in the context of the ND LTD and MetaArchive Cooperative, which are each committed to the long-term advancement of ETD programs in the United States. Our collaborative group will maintain and continue building on these documents and software tools in the course of future projects, as evidenced by the sustained efforts in previous years that have led to this proposal. In the course of this project we will build additional collaborative relationships with the USETDA, the TxETDA, and other emerging ETD professional associations. Additional details on long term maintenance of products are provided in the supporting documentation in *Appendix D. Future Plans*, and the Digital Products form.

We see this project as an essential catalytic step in fostering more sustainable and successful campus programs with the expertise and capabilities to manage all aspects of the extended lifecycle of ETDs.

LIST OF KEY PROJECT STAFF AND CONSULTANTS

1. *Dr. Martin Halbert* (UNT Dean of Libraries and Associate Professor)
2. *Dr. Katherine Skinner* (Educopia Institute Executive Director and MetaArchive Cooperative Manager)
3. *Gail McMillan* (VA Tech Director, Digital Library and Archives)
4. *Matt Schultz* (Educopia/MetaArchive Collaborative Services Librarian)
5. *Geneva Henry* (Rice University Executive Director, Center for Digital Scholarship)
6. *Dr. Bill Donovan* (Boston College Digital Preservation Manager & ETD Administrator)
7. *Xiaocan (Lucy) Wang* (Indiana State University Digital Repository Librarian)
8. *Dr. Patricia Hswe* (Penn State University Digital Collections Curator)
9. *Yan Han* (University of Arizona Associate Librarian for Research Services)
10. *Dr. Daniel Alemneh* (UNT Metadata Librarian)
11. *Shannon Stark* (UNT Strategic Projects Librarian)