

# **Preservation & Curation of ETD Research Data & Complex Digital Objects**

## **Narrative**

---

### **ASSESSMENT OF NEED**

Colleges and universities and the programs responsible for managing the submission, archiving, and dissemination of electronic theses & dissertations (ETDs) increasingly are concerned about administering and making available the research datasets and complex digital objects (multimedia files, software, etc.) that often accompany these scholarly works. These assets represent a rich source of information that can help to substantiate the dissertation and provide assurance that the foundation for replicable research is available for future researchers. These datasets and complex digital objects are also of great value and interest to libraries seeking to advance their missions in the digital age. ETD datasets and complex digital objects however, do pose curation and preservation challenges for institutions in terms of the viability of formats, file sizes, accessibility rights, dissemination pathways, and change-management issues. Recent reports, workshop evaluations, projects, surveys, and focus groups have indicated that U.S. colleges, universities and ETD/IR programs of all sizes are requesting and in need of *generalized yet adaptable guidance documentation, shared curation technologies, and corresponding training materials* to bridge the preservation and curation gap for these valued scholarly assets.

The Educopia Institute, in partnership with University of North Texas, and in concert with the Networked Digital Library of Theses and Dissertations (NDLTD), HBCU Alliance, bepress, ProQuest, and the libraries of Virginia Tech, University of Tennessee, Purdue, Carnegie Mellon, Oregon State, Penn State, Morehouse, University of Louisville, and Indiana State propose a two-year project to improve ETD policies and practices around research data and complex digital object management nationally. The project will answer the question: *How will institutions ensure the longevity and availability of ETD research data and complex digital objects (e.g., software, multimedia files) that comprise an integral component of student theses and dissertations?*

This project proposal has been informed by high-level investigations into data management needs and academic libraries' prospective roles in addressing them during a period of rapid change and strain due to funding mandates, resource shortages, and dynamically changing technologies. Over the last five years, the scholarly community has focused increasingly on "Big Data" curation and the preservation of federally funded research. Recent reports underscore the challenges involved in bridging the gaps between libraries and data producers, but also highlight the importance of cross-community collaborations to achieve the necessary scale and expertise to achieve data curation goals.<sup>1</sup> Recent reports have also identified a number of challenges that accompany today's requirements to better manage and curate research data. In a 2012 CLIR report, "The Problem of Data," Lori Jahnke, Andrew Asher, and Spencer Keralis deliberately address graduate researchers as a community of importance.<sup>2</sup> Graduate students are tomorrow's lead researchers; teaching them about research data curation now is key to the durability of future scholarship. Jahnke *et al*'s recommendations include:

- Emphasizing engagement with researchers and fostering dialog around identifying/building the appropriate tools for a particular project or set of projects rather than seeking a one-size-fits-all approach.
- Focusing educational or other training programs on early intervention in the researcher career path.
- Integrating data curation systems within an active research phase (as a backup and collaboration solution).

The authors argue that scholars need "archival skills to help them set priorities for data curation tasks and decide which data should be preserved," and that "...file formats, as well as the software and hardware platforms used to manage and manipulate data, tend to proliferate. Data preservation strategies not only must take into account these varied, proprietary, and non-standard data formats, but also must provide a real-time benefit for the scholar in meeting research goals." (p. 11) Student researchers and ETD/IR programs need *documentation, tools, and training* that encourage mutual knowledge growth and productivity.

---

<sup>1</sup> See e.g. the ARL report, "New Roles for New Times: Digital Curation for Preservation" (Skinner and Walters, 2011) and the CLIR report "Research Data Management: Principles, Practices, and Prospects." (Halbert, et. al., 2013). See also the Research Data Alliance's recent work on case studies and policies, including ["Data Publishing 2020: Four case statements"](#).

<sup>2</sup> See "The Problem of Data: Data Management and Curation Practices Among University Researchers" (Washington D.C., CLIR).

Recent data training efforts for graduate students underscore the importance of these recommendations and point to next steps. For example, the University of Massachusetts, Amherst Libraries' Data Working Group recently shared findings from their "Data Management Basics" workshop that documented the following:<sup>3</sup>

- "students are overwhelmed by the amount and variety of data they encounter in their research and express a desire to learn about effective techniques and tools to stay organized." (p. 182)
- "attendees reported that they didn't have a clear understanding of metadata and its role in data management practice." (p. 185)
- "attendees wanted information about concrete resources they could use in data management." (e.g., backup strategies, versioning, mapping data files, etc.) (p. 185)
- "attendees also wanted information about general online resources where they could learn more about data management." (p. 185)

Attendees further emphasized that in the future they would like this training to become much more discipline-specific and hands-on, requesting "Give more specific examples, maybe choose a single field with good data management practices and show a data management plan," and "Show us large files and talk about what makes a good vs. poorly organized data set." (p. 186) Any future *documentation, curation technologies, and training materials* for student research data and complex digital objects also need to be *generalized yet adaptable*.

Our project partners have pioneered efforts to provide student researchers with data management overviews:

- Purdue University Libraries has tested and documented approaches for outreach to campus research stakeholders. Their 2008 *e-Data Task Force* provides an early model of assessing library capacity and evaluating effective strategies for engagement. (Newton, Miller, and Bracke, 2011) Their *Introduction to Institutional Data Repositories Workshop* (Witt and Cragin, 2008) and recent research has helped to educate and equip student researchers with practical tools that promote fluency in data curation and preservation.
- The *Data Management Bootcamp for Graduate Students* workshop series, a joint program of Virginia Tech and four other Virginia universities, provides training to student researchers on data curation topics, including Understanding and Organizing Data; Formats and Transformation; Documentation and Metadata; Storage and Security; Data Protection; Rights and Access; and Preservation, Sharing and Licensing.
- The Network Digital Library of Theses & Dissertations (NDLTD) offered an initial Data Curation Workshop at ETD 2011 in South Africa that stressed the vital role students play in ensuring that their research data and other supplemental files have the necessary form, organization, and description.<sup>4</sup>

In addition, the University of Virginia hosts a Graduate Student Data Management Portal<sup>5</sup> that orients student researchers to the data curation lifecycle, and offers practical, well scoped, and subject-specific guidance on research data management as well as links to campus resources and recommended tools. These early efforts lay the groundwork for the development of resources that can help students and ETD programs at a national level.

The MetaArchive Cooperative hosted a 2014 focus group on current efforts addressing preservation and curation for ETD research data and complex digital objects. The focus group spotlighted the above-mentioned Data Management Bootcamp; the use and adaptation of the New England Collaborative Data Management Curriculum by the University of Tennessee, and a current graduate level data management course offered by Oregon State University.<sup>6</sup> The focus group highlighted challenge areas as requiring research and attention:

---

<sup>3</sup> See "Data Management Training for Graduate Students at a Large University" (Adamick, Reznick-Zellen, and Sheridan, 2013)

<sup>4</sup> See the *Data Curation Profiles Project* (<http://wiki.lib.psu.edu/display/dcp/Purdue+UIUC+Data+Curation+Profiles+Project>); *Data Management Bootcamp for Graduate Students*: <https://www.research.vt.edu/announcements/11-05-13/graduate-student-data-management-bootcamp>; NDLTD Data Curation Workshop <http://dl.cs.uct.ac.za/conferences/etd2011/workshops>

<sup>5</sup> See [http://pages.shanti.virginia.edu/SciDaC\\_Grad\\_Training/](http://pages.shanti.virginia.edu/SciDaC_Grad_Training/)

<sup>6</sup> See <http://library.umassmed.edu/necdmc/index> and <http://guides.library.oregonstate.edu/grad521syllabus>

- Documented strategies for working with faculty to ensure improved curation of graduate student research—e.g., embedding instruction in existing course curriculum
- Training on how to leverage library and other institutional staff and expertise to advocate for better preservation and curation practices early in the researcher’s career
- Providing clear and straightforward curation support and education for better data and content management—particularly with respect to metadata, file formats and version control, among other topics.

The programs above are exceptions, not the rule—institutions with higher-than-average resources have undertaken most of these early efforts. Others are not yet able to address data management adequately, and ETD/IR programs and student researchers seek guidance that can apply across many environments and fields. This project seeks to document and elucidate the persistent challenge areas and strategies and solutions in data curation for student research *and* to extend these undertakings to a broader range of institutions.

This project will coalesce, refine, and shift this knowledge accumulation out of its campus or regional contexts and build upon it to create more *generalized yet adaptable guidance documentation, shared curation technologies, and corresponding training materials* that can be adopted broadly by ETD/IR stakeholders.

### **IMLS Lifecycle Management of ETDs (<http://metaarchive.org/imls>): Project Assessment of Needs**

Through generous funding from IMLS, the University of North Texas, NDLTD, Educopia Institute, MetaArchive Cooperative, and six university libraries have researched and documented lifecycle curation for ETDs, producing deliverables that are now widely used by the academic community. The project inspired the deployment of a 2013 NDLTD survey to gauge the current state of ETD programs and their perceived needs that overwhelmingly demonstrated that ETD/IR programs report needing help with those programs through targeted guidance documentation, online materials, and workshops in that order of preference and priority.<sup>7</sup>

Through the work of the project and related NDLTD survey, we have determined a clear need for documentation, tools, and training materials aimed specifically at the *supplemental research data and complex digital objects* that often accompany the scholarly thesis. Repeatedly, project reviewers and advisors brought up the dense thicket of issues surrounding complex digital objects (e.g., software, multimedia files, digital art, and other material that sometimes is integral to the thesis or dissertation itself but that most campuses do not yet collect or preserve. Specifically, *ETD Guidance Documents for Lifecycle Management* reviewers requested additional targeted “how-to” documentation focused explicitly on file format issues, preservation metadata, and other digital object-specific curation tasks as a next crucial step in ETD program maturation. This feedback concurred with 2013 project workshop evaluations that identified the need for skills/knowledge in the management of research datasets and complex digital objects. The project’s Lifecycle Management Tools (currently undergoing final testing, packaging and dissemination) provide a strong foundation for helping *institutions* understand what curation technologies are available. This project will offer additional platform and implementation support to assist with testing and adoption in more localized settings where workflows are often highly institution-specific, and also with experimentation and learning by students who lack campus support.

### **Colleagues and Collaborators**

The DMPTool<sup>8</sup> service and the recent IMLS-funded DataFOUR and CRADLE<sup>9</sup> projects are related to this proposal and relationships between our teams will benefit all three initiatives. This project will not duplicate their advances but rather will benefit from and contribute to their outputs. DMPTool provides customizable frameworks that assist faculty to build data management plans for their federally funded projects. DataFOUR and CRADLE are aimed at addressing research data curation needs according to the broadest definition of the content genre and are weighted heavily towards faculty researchers and dedicated data curation professionals.

<sup>7</sup> The survey results are summarized on the project website: [http://metaarchive.org/imls/index.php/2013\\_NDLTD\\_Survey\\_Report](http://metaarchive.org/imls/index.php/2013_NDLTD_Survey_Report).

<sup>8</sup> See [www.dmp.cdlib.org](http://www.dmp.cdlib.org). PI (Skinner) is an *ex officio* Steering Committee member for DMPTool, and will bridge these efforts.

<sup>9</sup> See <http://imls.gwla.org/>, <http://sils.unc.edu/news/2013/tibbo-odum-imls-craddle>

Our proposed project's needs assessment has determined that *student researchers* in pursuit of their thesis or dissertation are amassing research data and other complex digital objects in unique ways that deserve integrated yet special attention. The ETD/IR programs supporting these student researchers include faculty, librarians, technologists, and vendors—audiences that will benefit from these integrated yet specialized resources.

## IMPACT

The project deliverables will have broad national impacts. These impacts will extend to a range of scholarly stakeholders including academic libraries, graduate schools, technology centers, and the authors and researchers that produce and make use of ETDs and their supplemental resources.

- ***Guidance Briefs:*** The team will document for ETD/IR stakeholders (including students, librarians, administrators, faculty, campus/library IT and others mentioned in the Diversity Plan) the full range of curation challenges that student researchers must navigate when preparing and depositing their supplemental content for long-term preservation and recommend practical strategies and policies that ETD/IR programs can promote to assist them. Based on findings from the Lifecycle Management of ETDs project and subsequent focus group work performed in preparation for this project, we have identified preliminary priority topics to address in short (2-4 page) “how-to” oriented Guidance Briefs on ETD dataset and complex digital object files: *Rationales and Motivations, Selection and Acquisition, Metadata, Change Management, File Formats, Data Structures, Documentation, IP Issues, and Access and Use*. These topics will be addressed with concrete guidance and recommended cross-discipline practices, and will feature illustrative discipline/domain-specific case studies. Prepared by the PIs and Project Manager and reviewed by the Advisors and Steering Committee, the documentation will help ETD/IR programs build and nurture supportive relationships with student researchers. These early interventions in graduate student researchers’ careers will help several groups—providing *students and faculty advisors* with much needed training in data management; *librarians* with a specific role in student and faculty workflows that increase the probability that the library will be an active participant in campus-based digital curation activities; and *researchers* with better access to foundational research materials. These briefs will also help student researchers understand how their approaches to data and content management impact credibility, replicable research, and general long-term accessibility—knowledge and skills that will impact the health of their careers for years to come.
- ***Curation Workbench:*** The team will develop a framework, workflow examples, and a pilot sandbox environment to facilitate hands-on learning and dissemination of proven preservation/curation technologies, which can then be downloaded and applied in localized ways for ETD research data and complex digital objects. Preliminary environmental scans, inventories, and gap analyses have surfaced a number of exemplars that inform this proposed deliverable.<sup>10</sup> These include the UVA Graduate Student Data Management Portal, as well as the Open Planets Foundation’s Plato Preservation Planning Tool, BitCurator, the Community-Owned digital Preservation Tool Registry (COPTR), and the Digital POWRR Project Tool Grid.<sup>11</sup> This project’s proposed *Curation Workbench* will produce an adaptable and updatable framework to facilitate adoption or retirement of specific modular technologies over time, making it possible for ETD/IR programs to identify (through workflow examples, not a sprawling registry), test (through a sandbox environment), and implement (locally, with installation documentation) the technologies best suited for designated stakeholders. The *Curation Workbench* also aims to reduce the barriers of intimidation for student researchers (and others) through providing clear documentation and a testbed environment. Modules of preliminary priority to include in these workflows and sandbox include file format recognition and migration tools, preservation packaging and ingest tools, preservation metadata tools, content versioning support, and file renaming tools. Because the landscape is rapidly changing, the project will kick off with an

<sup>10</sup> See Appendix B: Curation Workbench Environmental Scan & Gap Analysis included with this application.

<sup>11</sup> See: <http://www.ifs.tuwien.ac.at/dp/plato/>; <http://www.bitcurator.net/>; <http://coptr.digipres.org/>; <http://digitalpowrr.niu.edu/tool-grid/>

additional set of environmental scans, inventories, and gap analyses to further refine the scope and inclusion of relevant tools. A functional requirements analysis will also be performed to ensure proper fit-to-purpose for the stakeholder communities. The project team will work closely with the BitCurator and CRADLE teams (through Cal Lee and Jon Crabtree) to develop specifications and a pilot instance of a sandbox environment that will allow students, technologists, librarians, and advisors to explore options in a hands-on manner. This plan was influenced through an ANADP II action session (Barcelona, 2013) hosted by Lee.<sup>12</sup>

- *Training Workshop:* The team will create and deploy a training workshop that encourages dialog and interaction across the various stakeholders and promotes measurable learning improvements, building upon the Guidance Briefs and Curation Workbench. With built-in interactive exercises and hands-on demonstrations, this modular training workshop will focus on orientation and instructions for adopting and using the main project deliverables in diverse local settings. It will build modules for tailored audiences, including ETD/IR programs, students, librarians, advisors, IT, or some combination of these. This approach to deploying such educational workshops has proven successful in fulfilling stated learning objectives through models such as the Library of Congress's Digital Preservation Outreach & Education Program (DPOE) and in the "Lifecycle Management of ETDs" project's workshops. The workshop, designed and taught by Skinner, Schultz, and Krabbenhoef with guidance from the Steering Committee and Advisors will be geared toward promoting on-going, pragmatic, and sustainable support to all of the targeted stakeholders involved in the ETD arena. Learning improvements will be measured through on-site pre- and post-workshop evaluations and a follow-up survey 1-month after the workshop. These will be designed and conducted with the team's Evaluator to ensure demonstrable success and measurable benefits.

---

## PROJECT DESIGN AND EVALUATION

The project is designed to gather the information necessary to produce three well-scoped and achievable deliverables and to measure their impact on the stakeholder community. The project's research methods, evaluation approaches, and targeted outcomes build upon the design and evaluation approaches utilized in the highly successful 2011 IMLS-funded ETD Lifecycle Management Project. That project conducted rigorous surveys, interviews, and focus groups, as well as environmental scans, inventories and gap analyses—all guided by an expert Steering Committee—to scope and construct all project deliverables. In terms of evaluation, the project deployed two phases of extensive review and improvements that consulted both partners and stakeholder communities for feedback. Feedback cycles helped to expand and enrich the deliverables' content, form, and dissemination and demonstrated substantive and significant improvements in knowledge and skills.<sup>13</sup>

### Research Methodology and Questions

As evidenced in the needs assessment, several recent community studies, such as those undertaken by RDA, CLIR, DLF, and IMLS, have sought to identify the scope and definition of research data and "Big Data" and the needs of faculty and professional researchers. Far fewer have addressed student researcher and early-career researcher needs in their creation, accumulation, organization, storage, and long-term archiving of data and complex supplementary content. To document needs and practices across all tiers of higher education, including smaller, less-endowed institutions, the project will conduct descriptive studies (surveys, interviews, and focus groups) and technical analyses (environmental scan, inventory, gap analysis and functional requirements). There are several key questions and information objectives that drive our research in this project.

- *Guidance Briefs:* What curation/preservation challenges specifically pertain to student research data and complex digital objects? How may institutions use internal resources and partnerships to address these challenges? What document form(s) will ensure the greatest impact on the broadest range of stakeholders?

---

<sup>12</sup> See <http://educopia.org/events/ANADPII/program>.

<sup>13</sup> See [http://metaarchive.org/imls/index.php/Lifecycle Management of ETDs: Impacts and Outcomes](http://metaarchive.org/imls/index.php/Lifecycle_Management_of_ETDs:_Impacts_and_Outcomes).

- *Curation Workbench:* Which existing curation/preservation technologies are most relevant to stakeholders? How do students and curators identify and implement appropriate curation/preservation technologies? Does access to experimentation (sandbox environment) and documented workflows increase implementation?
- *Training Workshop:* How may institutions develop policies, programs, and services to support curation/preservation of student datasets and complex digital objects? What technologies are most important to demonstrate for students, for ETD/IR programs, for libraries, and for IT professionals? What training can be generalized and what needs to be tailored for specific stakeholder audiences?

The project will deploy the following methods to refine the scope of the stated deliverables at project launch.

- *Surveys, Focus Groups, and Interviews:* We will deploy surveys to gather deep information regarding current stakeholder community needs at launch. Feedback from these surveys will inform the content, form, and dissemination preferences for each deliverable. Surveys will identify stakeholder perceptions and recommendations for achieving institutional adoption and long-term usage of project deliverables. After analyzing findings, we will host focus groups and interviews with representatives of key stakeholder groups to test our survey-based hypotheses and verify content/form/access preferences for the deliverables.
- *Environmental Scan, Inventory & Gap Analysis:* In addition to the research we have undertaken prior to this proposal, we will also perform a second detailed environmental scan, inventory and gap analysis. The current pace of technological change demands this additional environmental scan to refine the scope of modules we consider for the workflows supporting curation of ETD research data and complex digital objects, and workbench platforms (e.g., Plato, BitCurator) and sandbox environments.
- *Functional Requirements Analysis:* Based upon the environmental scan, inventory & gap analysis we will refine the functional requirements (e.g., format recognition and migration tools, preservation packaging and ingest tools, preservation metadata tools, etc.) within the proposed online workspace, as well as corresponding use cases and usage documentation for the intended stakeholder groups. These functional requirements will guide alpha and beta stages of development, partner testing, and a public review period.

## Evaluation Methods and Outcome Assessments

The project includes structural and strategic evaluation and outcome measurement activities. These include:

- *Evaluator:* We will engage an Evaluator Consultant (Drummond) to advise on data gathering methods (surveys, focus groups, and interviews) and design implementation evaluations. The Evaluator will help design the partner site and public review evaluation processes for all deliverables. In preparation for the in-person Advisory Group Meeting, she will work with the project team to review, prioritize, and develop strategies to address evaluation findings. She will assist with measuring and reporting project impacts.
- *Project Steering Committee Meetings:* The project will use monthly project calls to analyze the findings from our surveys, focus groups, interviews, environmental scans, inventories, gap analyses and functional requirements analyses. The Steering Committee (including reps from universities, libraries, and service providers) will evaluate and test the draft deliverables and recommend improvements prior to public release.
- *Project Advisory Group Meetings:* The project has also enlisted a select group of experts in the areas of copyright (Dwayne Buttler, Univ. of Louisville), research data (Michael Witt, Purdue), policy (Kathleen Shearer, COAR), ETDs (Gail McMillan, NDLTD), preservation/curation technology (Christopher “Cal” Lee, UNC), metadata (Amy Barton, Purdue), and academic libraries (Tyler Walters, VA Tech). These Advisors will meet with the Project Personnel on a bi-monthly and as-needed basis to evaluate and inform the project. They will also evaluate all deliverables and outcomes at an in-person Advisory Group meeting.
- *Partner Site Evaluations:* Prior to public review, the team will work with the Steering Committee to deploy and test the draft deliverables among their campus stakeholder communities. Using instruments designed to provide qualitative and quantitative feedback, each deliverable will be evaluated.

- *Public Evaluations:* After partner site evaluations conclude, each deliverable will be refined and released for a four-month public review period. The public reviewers will provide qualitative and quantitative data on fit-to-purpose, effectiveness, and areas for needed improvement. Feedback will guide refinements.

---

## DIVERSITY PLAN

The stakeholder communities interested in managing ETD research data and complex digital objects occupy distinct professional roles and responsibilities. These stakeholder communities and their parent campuses also represent diverse demographics, ethnicities, finances, and perspectives that will be studied in our research methods and deliverables. Target communities are represented via the Project Steering Committee and Advisory Group and are deeply engaged in the project's research. The descriptive studies will be conducted to identify individual and organizational demographics and particularly note perspectives of traditionally underserved groups, including non-ARL institutions, women's colleges, and historically black colleges and universities.

**1. Student Researchers:** The project will study both graduate and undergraduate attitudes and approaches to preserving and curating supplemental ETD content, and address outstanding challenges/barriers to better data and content management across disciplinary domains/methods. The project team includes multiple PhD graduates who have recently grappled with these issues (e.g., PI Katherine Skinner; Advisor Tyler Walters).

**2. ETD/IR Program Administrators:** These decision-makers evaluate and shape policy, and allocate institutional resources to manage the institution's scholarly outputs. We will document how this group can bridge curation gaps with student researchers and collaborate around shared technologies to ensure ETD data files' long-term accessibility. The project team includes an ETD Program Expert (Gail McMillan, VA Tech).

**3. Library Administrators:** Deans/directors, associate deans, and department heads often facilitate the creation of ETD/IR programs and ensure their fiscal support and sustainability. The project will gauge the level of importance and strategic priority these stakeholders place on collection/management of ETD research data and complex digital objects. The team includes library deans (Martin Halbert, UNT and Tyler Walters, VA Tech).

**4. Graduate School Administrators:** These stakeholders include the graduate council, graduate faculty, and the deans and associate deans, as well as the graduate school staff that work closely with the ETD/IR program—namely student service officers. The project will engage these stakeholders to help define their specific roles helping students who are creating and curating research data and other complex digital objects. The project Advisors includes a faculty member who regularly chairs dissertation committees (Cal Lee, UNC).

**5. Institutional Administrators:** This group includes top-level decision-makers such as the university president, provost, and chief information officer. They are not directly involved with the day-to-day affairs of student research or the institution's ETD/IR programs or research support services but are critical to the oversight and on-going support of these activities. The project will engage this group of high-level stakeholders to gauge what incentives they see for encouraging curation of ETD/IR research data and supplemental content.

**6. Campus Research Centers/Support Services:** These dedicated hubs of professionals, infrastructure, and services often assist in the creation and curation phases of student research. The project has recruited a Research Data Specialist (Michael Witt, Purdue) as an Advisor, and will seek to understand how this stakeholder community can be enlisted and leveraged to effectively align with student researchers and ETD/IR programs .

**7. General Counsel/Copyright Librarians:** Legal issues play a crucial role in determining how research is conducted and how research outputs are documented and shared—including determining permissions to archive and provide access to ETDs and supplemental data and content. They assist researchers with guidance on intellectual property and copyright concerns, and help ETD/IR programs set policy regarding embargoes. The project team includes a copyright specialist (Dwayne Buttler, Univ. of Louisville) as a Project Advisor and will consult with stakeholders to provide clear guidance on copyright and IP concerns.

**8. Librarians & Archivists:** Library staff provide expert guidance and practical support to ETD/IR programs as they seek to establish and offer services for depositing, indexing, archiving, and providing or restricting access to ETDs and their supplemental content. The Steering Committee's librarians and archivists (various) and Advisors (Amy Barton, Purdue) and our research (surveys, interviews, focus groups) will help us identify the channels by which these stakeholders contribute toward curating supplemental ETD data and content files.

**9. Library/Campus IT:** This project is especially interested in the group of stakeholders represented by the library and/or campus IT managers and staff. These technologists play a vital role in shaping the ETD/IR technology/infrastructure that provides the foundation to support the preservation and curation of ETD research data and complex digital objects. In addition to the findings that will emerge from targeting this group in the project's research, the project has also enlisted multiple library technologists to serve as Project Advisors.

**10. ETD/IR Professional Associations:** In and beyond the U.S., international, national and regional organizations provide networks and guidance to ETD/IR programs. These stakeholders promote education, training, networking events, and make available resources that assist ETD/IR program managers and related staff in their professional efforts. The project has enlisted expertise from the Confederation of Open Access Repositories ((Kathleen Shearer, COAR) and the NDLTD (Gail McMillan, VA Tech) on the Project Advisory Group, and received written support from the NDLTD, USetdA, and TxETDA.

**11. ETD Service Providers:** Commercial and non-profit companies serve college and university ETD/IR programs, offering products and services such as archiving and cataloging of ETDs, and offering subscriptions to extensive databases of ETDs and their related content. The project has incorporated a leading ETD service provider—ProQuest—on its project Steering Committee (Austin McClean), and will actively solicit additional service providers' perspectives to incorporate their perspectives and supporting use cases.

**12. Institutional Repository Services:** IR software systems support the archiving, management, and dissemination of ETDs and their related supplemental content (e.g., DSpace, CONTENTdm, bepress, Fedora, and ArchivalWare). The project Steering Committee includes a bepress representative (Eli Windchey) to ensure that the project deliverables have adoption potential for this important cohort of stakeholders.

---

## PROJECT RESOURCES AND PLAN

This project requests \$250,000, matched with \$68,987 in cost share, to build on successful collaborative research undertaken by the Educopia Institute/MetaArchive Cooperative, and the NDLTD. Our team has a strong sense of the next steps needed to advance ETD/IR program resources nationally, and has proposed this project to move this work forward. The following are descriptions of project personnel and areas of expertise.

**Dr. Katherine Skinner** (PI; Educopia Institute Executive Director) will act as PI and convener of the project. Skinner will devote 10% of her time as cost-match to the project. She has co-PIed the Lifecycle Management of ETDs project. She will coordinate steering committee work on the Training Workshop.

**Matt Schultz** (PI; MetaArchive Cooperative Program Manager) will act as PI and convener of the Advisory Board. Schultz will devote 10% of his time as cost-match to the project. He will coordinate steering committee work on the Curation Workbench and supervise the Project Manager.

**Dr. Martin Halbert** (Co-PI, University of North Texas; UNT Dean of Libraries and Associate Professor) will act as a co-PI, devoting 7% of his time as cost-match to the project. The UNT Libraries recently hosted the ETD Lifecycle Management project. He will coordinate steering committee work on the Guidance Briefs.

**Nick Krabbenhoeft** (Project Manager; Educopia Project Manager): Krabbenhoeft will act as *project manager* with 100% of his time devoted to this role. He will supervise the programmer and assist with technical writing. Krabbenhoeft works on targeted MetaArchive digital preservation contracts on a regular basis.

**Stephen Eisenhauer** (Project Programmer, 65%): Eisenhower is an experienced digital library technologist,

familiar with the concepts of micro-services in the context of lifecycle management of digital assets. He is based at UNT, and has served as the programmer on the IMLS-funded Lifecycle Management of ETDs project.

**Steering Committee:** Members will assist with research, deliverables scoping, and evaluation of deliverables.

**Advisory Group:** Members will oversee the project and assist in evaluation of deliverables and outcomes.

**Christina Drummond (Evaluator):** The evaluator will spend 50 hours consulting with the project team regarding survey/focus group design and analysis, deliverable public review structure and questions, and outcomes evaluation for the project overall. Drummond provides strong experience and an external perspective.

## **Workplan**

The project begins October 1, 2014 and concludes September 30, 2016. *Additional details of the work plan are provided in the Schedule of Completion GANTT chart.* The following provides a high-level summary.

### **Preparation Phase (October 2014 – November 2014)**

*Activities:* The first two months will facilitate communication between project partners and orient them to the project's workspaces, schedule, and expected workflows. We will host kick-off meetings for the Project Personnel, Advisory Committee, and Steering Committee, and will refine plans for the Research & Development Phase with their input. Project Personnel will document an Outreach Plan and set preliminary dates for the in-person Advisory Group meeting. *Deliverables:* Research & Evaluation Plans and Outreach Plan.

### **Research & Development Phase (December 2014 – February 2016)**

*Activities:* Building upon our in-depth pre-proposal needs assessment, the project will administer a second round of surveys, focus groups, and interviews to further clarify the needs of the stakeholders (see Diversity Plan), and refine the *Guidance Briefs* scope and topics. Partners will identify recipients/participants for these activities and project personnel will contact them through listservs/emails. Contact information will be stored during the project period in a secure database for reuse in project evaluation. After two months of research, we will develop drafts of the *Guidance Briefs* and review and refine them internally over an 11-month period. Simultaneously, we will conduct a thorough environmental scan to update and complement our pre-proposal research on the *Curation Workbench*. We will document functional requirements to guide the alpha and beta development phases and internal testing of the tools over a 15-month period. The project partners will plan the *Training Workshop*. All deliverables will undergo rigorous public reviews. *Deliverables:* launch of packaged *Guidance Briefs* and *Curation Workbench* for public review, and a packaged pilot *Training Workshop*.

### **Evaluation & Improvement Phase (February 2016 – July 2016)**

*Activities:* Using our contacts database and major listservs, the *Guidance Briefs* and *Curation Workbench* will be distributed for four-month public reviews. Reviewer feedback will be structured through the use of review questions and anonymized by the Project Manager before being distributed to Steering Committee members and project personnel. The Steering Committee and Personnel will complete the final revisions over a 3-month period. The PIs will deliver a finalized version of the *Training Workshop* in conjunction with a major conference of relevance to the stakeholder communities. An in-person Advisory Group meeting will also be held to review and refine all project deliverables and assess project outcomes. *Deliverables:* Final versions of the *Guidance Briefs*, *Curation Workbench*, and *Training Workshop* ready for distribution.

### **Reporting & Completion Phase (August 2016 – September 2016)**

*Activities:* The project partners will report significant findings during each project phase. This final reporting phase focuses on the packaging and dissemination of project deliverables using various distribution channels, including listservs, the project contacts database, and presentations at major events serving the ETD/IR, graduate schools, and research data communities. The packaging and dissemination of project materials will be carried out in accordance with the Sustainability Plan (see below) and tracked through web analytics. With

input from the Evaluator, the final report will include clear outcomes evaluation. *Deliverables:* Archived deliverables, presentations, ongoing analytics on project deliverables downloads, final report.

---

## **COMMUNICATION PLAN**

There will be multiple opportunities and communication channels for relevant stakeholder groups to engage with this project and its deliverables. The goal in providing these opportunities and communication channels is to invite on-going feedback, raise awareness, and support the project's stated impact objectives. Specific communication activities include: 1) *Project Survey:* Surveys conducted early in the first year will gather real-time baseline information about the current state of preservation and curation training and resources for student researchers regarding supplemental ETD data and other content files. 2) *Project Focus Groups:* Focus groups involving representatives from the stakeholder groups will be carried out early in year one to identify both the common challenges and proven solutions to bridging gaps in preservation/curation and stakeholder cooperation/collaboration. 3) *Project Website:* A project website established by Educopia Institute will serve as a reliable, on-going resource for the various stakeholder groups and the public at large to obtain general information, progress updates, and deliverables. We will also work with our partners, the NDLTD, to circulate this information through its website. 4) *Published Papers & Presentations:* The extended project team will hold an outreach-planning meeting in the first months of the project to identify conference presentation opportunities. Presentations will be a critical channel for communicating research findings and describing the project deliverables as they take shape. Papers and presentation slides for any selected and attended events will be archived and made publicly available on the project website. 5) *In-Person Advisory Group Meeting:* In year two, the Advisors, Evaluator, and Personnel will meet to review and refine all project deliverables. This meeting will be held in conjunction with a conference event of relevance to the stakeholder communities. As such the project team will take advantage of this opportunity to engage with those communities. 6) *Training Workshop:* In year two there will be both a pilot and a final training workshop delivered to the stakeholder community to provide orientation and instructions for adopting and using the main project deliverables. These workshops will be held in conjunction with a conference event of relevance to the stakeholder communities.

---

## **SUSTAINABILITY PLAN**

The project deliverables will be among the first resources of their kind for ETD/IR programs and student researchers. As such it is critical that they are easily accessible and appropriately documented and licensed to ensure that the stakeholder community can readily use and adapt them going forward. The main project website, hosted by Educopia Institute, will promote all review and final versions of the project deliverables to the main page—the site will be open to archiving by the WayBack Machine and be outfitted with Google Analytics. All project background, reports and summary findings will also be just a click away from this main page. The second element of sustainability support will exist within and around the deliverables themselves. Each will be open source and Creative-Commons licensed to promote use and reuse with appropriate citation. Each will also be documented with tips and instructions for use, enhancement, and adaptability by stakeholder user groups. This will be especially important for the *Training Workshop*, which aims to communicate *generalized yet adaptable* use cases for the *Guidance Briefs* and *Curation Workbench*. Finally, the project website and deliverables will be listed as recommended resources for ETD/IR programs and student researchers via the NDLTD website. Efforts will also be made throughout the project to work with any interested institutions that would like to host and deploy customized versions of the project deliverables. 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 NDLTD and Educopia/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.