

*University of North Carolina at Chapel Hill – BitCurator.edu*

The University of North Carolina at Chapel Hill School of Information and Library Science (UNC SILS), in collaboration with the Educopia Institute, the BitCurator Consortium and the Council of State Archivists (CoSA), requests \$499,664 for a three-year Laura Bush 21<sup>st</sup> Century Librarian Program (LB21) research grant (7/1/2018 – 6/30/2021) to study and advance the adoption of digital forensics tools and methods in libraries and archives through professional education efforts. The audience for this project is providers of professional education for information professionals (both university-based and continuing professional education).

Essential to advancing the **National Digital Platform** is the capacity of libraries and archives to acquire, manage and provide access to born-digital materials. There have been significant recent advances in applying digital forensics methods in libraries and archives to address this need. Measures include creating disk images to ensure the completeness, authenticity and availability of data; using cryptographic hashes for de-duplication and to identify, verify and authenticate materials; generating metadata to document the chain of custody; identifying and extracting contextual information; finding and locating sensitive information for further review, redaction or removal; building access environments that allow users to access data from disks using a web browser or emulation platform; and generating specialized reports to characterize contents of collections.

Education is an essential component of advancing new professional practices, and there have been important advances in digital forensics education for library and archives professionals. The BitCurator.edu project will engage in research to investigate and support graduate-level LIS/preservation programs to implement educational offerings that make use of digital forensics tools and methods. This project will address two primary research questions:

RQ1: What are the primary institutional and technological factors that influence adoption of digital forensics tools and methods in LIS classes in different educational settings?

RQ2: What are the most viable mechanisms for sustaining collaboration among LIS programs on the adoption of digital forensics tools and methods?

In addressing RQ1, the project team will work with members of the Advisory Board to test and implement several models for administering digital forensics education. The project team will share guidance documents from these activities, which will serve as resources for education efforts elsewhere. RQ2 will be addressed through engagement with the BCC, Advisory Board and PEP on sustainability models, including (but not limited to) development of a new educational membership category for the BCC.

The Advisory Board represents a diverse set of schools: Catholic University, Indiana University, New York University (Moving Image Archiving and Preservation Program), San Jose State University, University of Illinois, University of Maryland, University of Michigan, University of Texas, Wayne State University. In addition to the project Advisory Board, we will also draw on a Professional Experts Panel (PEP) of individuals from the Council of State Archivists (CoSA), Digital Library Federation (DLF), The Lyndhurst Group, LYRASIS, Massachusetts Institute of Technology (MIT), National Museum of American History, and the Southeastern Museums Conference (SEMC), who will provide input and guidance on all aspects of the project through periodic conference calls.

This project will dramatically advance the **National Digital Platform** by 1) producing generalizable findings about factors influencing various educational implementation approaches and strategies for addressing them, 2) building the capacity of educational programs across the country to better prepare their students to effectively manage, preserve and provide access to born-digital materials, and 3) providing exemplars and guidance on experiential learning opportunities that can serve as models for other institutions. We expect the project to achieve the following measurable outcomes: 1) strengthen information sharing across partners, 2) testing of learning objects by partners, 3) adoption of learning objects by institutions beyond the partners, and 3) increase professionals' knowledge of and fluency in digital forensics instruction.

## 1. Statement of Broad Need

A rapidly growing body of materials with significant cultural value are “born digital.” Libraries and archives are increasingly called upon to move born-digital materials that are stored on removable media (e.g. floppy disks, flash drives, CD-ROMs, hard drives) into more sustainable preservation environments. This is a relatively new function, and one that is rife with technical and ethical challenges, and most librarians and archivists have limited access to training that addresses this crucial activity.

Essential to advancing the **National Digital Platform** is the capacity of libraries and archives to acquire, manage and provide access to such born-digital materials. Information professionals must be prepared to extract digital materials from removable media in ways that reflect the rich metadata and ensure the integrity of the materials. They must also support and mediate appropriate access: allowing users to make sense of materials and understand their context, while also preventing inadvertent disclosure of sensitive data.

There has been a significant shift in recent years toward the adoption of digital forensics tools and methods by libraries and archives in order to meet these goals. A small number of librarians and archivists have forged pathways for creating disk images to ensure the completeness, authenticity and availability of data; using cryptographic hashes for de-duplication and to identify, verify and authenticate materials; generating metadata to document the chain of custody; identifying and extracting contextual information; finding and locating sensitive information for further review, redaction or removal; building access environments that allow users to access data from disks using a web browser or emulation platform; and generating specialized reports to characterize contents of collections.

This work has been advanced by development and distribution of free and open-source software designed to support work in collecting institutions. The BitCurator project series (2011-2018), funded the Andrew W. Mellon Foundation, has developed the open-source BitCurator Environment (which is used actively by collecting institutions around the world to carry out the tasks listed above) and has developed, documented and disseminated technologies that simplify and enhance access to born-digital data. The BitCurator Consortium (BCC) is an independent self-sustaining membership organization responsible for maintenance of the products of the BitCurator projects, as well as advancing the community of professionals who use them.

As valuable as digital forensics tools and methods can be in advancing digital curation goals, they cannot be adopted by professionals without relevant knowledge and skills. Those involved in this work must understand the nature of the underlying technology (Lee, 2012), repository requirements (Woods, Lee and Garfinkel, 2011), and metadata considerations (Chassanoff, Woods and Lee, 2017). There have been notable advances in digital forensics education for library and archives professionals over the last few years.<sup>1</sup> However, limited education offerings continue to be a barrier to adoption of digital forensics in libraries and archives, despite advances in both tools and models for instruction. Relatively few LIS graduate education programs currently have the in-house capacity to thoroughly expose their students to these practices. The number of instructors providing continuing education is still quite limited; for several years, BitCurator.edu PI Lee taught the majority of them. Expanding both the pool of instructors and training opportunities for students and professionals in

---

<sup>1</sup> UNC SILS has developed in-depth course offerings and a state-of-the-art digital forensics lab for research and teaching. Several other LIS programs (represented on the project Advisory Board) have also introduced hands-on digital forensics elements into their courses. On the continuing education front, a set of digital forensics classes are now offered through the Digital Archives Specialist (DAS) curriculum of the SAA, as well as modules in CoSA’s State Electronic Records Initiative (SERI) Institutes, and various specialized workshops. These initial offerings are limited to a small number of institutions and professional associations.

these fields requires that we design and implement simplified pathways to gaining expertise, including the hands-on, lab-oriented environments in which we can teach digital forensics.

### 1.1. Previous Work - Digital Forensics Software for Libraries and Archives

The BitCurator project (2011-2014), funded by the Andrew W. Mellon Foundation, was a joint effort led by the School of Information and Library Science at the University of North Carolina, Chapel Hill (SILS) and the Maryland Institute for Technology in the Humanities (MITH) to develop an open-source system for collecting professionals that incorporates the functionality of many digital forensics tools. BitCurator addressed two fundamental needs for collecting institutions absent from software designed for the digital forensics industry: incorporation into the workflow of archives/library ingest and collection management environments, and provision of public access to the data. BitCurator defined and tested support for a digital curation workflow that begins at the point of encountering holdings that reside on removable media—either new acquisitions or materials that are within a repository’s existing holdings—and extends to the point of interaction with an end user.

The BitCurator Environment is built on a stack of free and open source digital forensics tools and associated software components, modified and packaged for increased accessibility and functionality for collecting institutions. The BitCurator software is freely distributed under an open source license. It can be installed as a Linux environment; run as a virtual machine on top of most contemporary operating systems; or run as individual software tools, packages, support scripts, and documentation. Tools in the BitCurator Environment – both those produced by the project team and those from third-party developers – help advance core digital curation activities, including (but not limited to):

- Reducing the risk of inadvertent changes to content through software-based write-blocking
- Creating authentic copies of content through disk imaging and cryptographic hashing
- Mounting forensically packaged disk images to view & export contents
- Reflecting original order of materials through capture of file system metadata
- Establishing trustworthy chains of custody through documentation of curatorial actions (log files, PREMIS records)
- Generating reports that characterize the contents of disks and directories
- Identifying and documenting duplicate files
- Discovering and exposing associated contextual information
- Identifying sensitive information to be filtered, redacted, or masked in appropriate ways
- Exporting contents of disks and directories for inclusion in Archival Information Packages and Dissemination Information Packages

Because no registration is required to download or use the software, it is not possible to get an exact count of BitCurator users, but there are many indicators of a thriving user base as well as demand for staff who know how to use the software. Numerous institutions have cited familiarity with the BitCurator environment in recent job postings, including the Rockefeller Archive Center, Harvard University, University of Virginia, Drexel University, University of Michigan, Cornell University, the Massachusetts Institute of Technology, and Tufts University. At the time of writing, the BitCurator Users group includes 216 individual members and an archive of nearly 200 discussions. The BitCurator Twitter account (@BitCurator) has more than 1800 followers, and is used to promote project activities, inform users of new software releases, and answer community questions.

The BitCurator Access project (2014-2016), also funded by the Andrew W. Mellon Foundation, developed open-source software that support access to disk images through three approaches: building tools to support web-based services (Misra, Lee and Woods, 2014), enabling the export of file systems and associated metadata, and the use of emulation environments (Woods et al, 2015). The BitCurator Access project also developed a

tool to redact files, file system metadata, and targeted bitstreams within disks or directories. BitCurator Access focused on approaches to simplify access to raw and forensically-packaged disk images; allowing collecting institutions to provide access environments that reflect as closely as possible the original order and environmental context of these materials. The use of forensic technologies allows for detailed metadata to be generated reflecting the provenance of the materials, the exact nature of the file-level items they contain, and the metadata associated with both file-level items and data not observed within the file system (but still accessible within the original materials).

One of the primary motivations for using the BitCurator and BitCurator Access software is to capture and provide access to contextual information. For example, the original filesystem attributes associated with files (e.g. directory paths, timestamps) can be essential to understanding their provenance and original order. However, there are many other types of contextual information that can be vital to making sense and meaningful use of digital objects (Lee, 2011). In a study of archives reference questions, Duff and Johnson (2011) found that most information requests were based on “proper names, dates, places, subject, form, and, occasionally, events when composing their information request.” In their study of genealogists, Duff and Johnson (2003) identified information seeking practices focused primarily on names, places and time periods.

The BitCurator NLP project (2016-2018), also funded by the Andrew W. Mellon Foundation, has been developing software for collecting institutions to extract, analyze, and produce reports on features of interest in text extracted from born-digital materials. The project is adapting existing natural language processing (NLP) software to identify and report on items likely to be relevant to ongoing preservation, information organization, and access, including entities (e.g. persons, places, and organizations), potential relationships among entities, and topic models to provide insight into how concepts are naturally clustered within the documents.

The BitCurator software is ideally suited for use in educational settings. First, it is free and open-source, so instructors and students can download and use it without concerns about paying or negotiating licenses. Second, the software is built on top of an entire operating system (Ubuntu Linux), which can be run within a virtual machine, making it easy to run on any Windows or Mac computer with at least 8 GB of memory. Third, the software is well-documented, with a Quickstart Guide, screencasts, and various other guidance documents. Fourth, the modular design allows educators to adopt only a subset of the tools (rather than the entire environment) as appropriate to their needs. Finally, the software is maintained and supported by the membership-driven BitCurator Consortium, administered by Educopia.

## 1.2. Previous Work - Digital Forensics Education for Librarians and Archivists

The field of digital forensics focuses primarily on criminal investigations and civil law suits. This work became particularly important in the 1980s, as many people began using personal computers and thus had the potential to commit crimes or other improprieties using them. So the process of professionalizing digital forensics work began only a few decades ago (Garfinkel, 2013). Recognizing digital forensics as a distinct profession or discipline is an ongoing challenge and subject of debate (Irons, Stephens and Ferguson, 2009; Losavio, 2016). The system of education for digital forensics reflects this incipient status of the profession. Systematic attention to digital forensics education began even more recently (Lang, 2014). There are relatively few courses of study in universities (Liu, 2016). Many practitioners in the field receive instruction from either government law enforcement or intelligence agencies or vendors of proprietary software. The former is not available to those outside of the criminal justice system; the latter can be prohibitively expensive and is tied to the use of software that is very expensive and difficult (or impossible) to customize.

The past decade has seen various efforts to address digital forensics education for librarians and archives.<sup>2</sup> UNC SILS has been one of the important centers of this activity. Lee developed a half-day workshop on this topic at the IS&T Archiving Conference in May 2009 and a full-day workshop for staff of UNC Libraries in August of the same year. From June 2010 to June 2011, Lee served as the PI for the Digital Acquisition Learning Laboratory (DALL) project, which was funded by the Andrew W. Mellon Foundation. The project team built expertise and infrastructure within the school to administer novel educational exercises. They also issued a white paper with lessons learned and examples of instructional materials (Lee and Woods, 2011). Lee and Kam Woods developed a new special topics course at SILS, Acquiring Information from Digital Storage Media (INLS 490-141), which they offered in Spring 2011; they have offered this dedicated digital forensics class at SILS several more times, and it is now a regular offering of the school (INLS 561: Digital Forensics for the Curation of Digital Collections). Other SILS courses that have included digital forensics materials and activities are Electronic Records Management (since Spring 2011), and Understanding Information Technology for Managing Digital Collections (INLS 465), which was introduced in Fall 2008 as an outgrowth of the DigCCurr (Digital Curation Curriculum) project funded by the IMLS (Lee, Tibbo and Schaefer, 2007).

The Advisory Board of BitCurator.edu includes professors and institutions across the US that have begun offering digital forensics education in recent years.<sup>3</sup> It also includes those who have developed and administered digital forensics continuing education for LAM professionals. For example, Lee incorporated core digital forensics exercises into the DigCCurr II Professional Institute, a week-long continuing education course offered annually in Chapel Hill (and once in Copenhagen) 2009-2016, as well as the State Electronic Records Initiative (SERI) Institute which was administered in Indianapolis, Indiana on July 8-12, 2013. In 2012, Lee began administering a series of Digital Forensics for Archivists classes for the Society of American Archivists as part of the DAS (Digital Archives Specialization) certification<sup>4</sup>; what began as a single one-day class is now a three-day series (one day of fundamentals and two days of advanced instruction). In 2010-2014 and then again starting in 2017, Matthew Kirschenbaum and Naomi Nelson have co-taught a course at the University of Virginia's Rare Book School (RBS); Kirschenbaum has several sessions dedicated to digital forensics as part of this curriculum.<sup>5</sup> The FIDO project also ran a training event called "Applying Digital Forensic techniques to AIM [Archives and Information Management]" in 2011 in London.<sup>6</sup>

## 2. Project Design

The University of North Carolina at Chapel Hill School of Information and Library Science (UNC SILS), in collaboration with the Educopia Institute, the BitCurator Consortium and the Council of State Archivists (CoSA), requests \$499,664 for a three-year Laura Bush 21<sup>st</sup> Century Librarian Program (LB21) research grant (7/1/2018 – 6/30/2021) to study and advance the adoption of digital forensics tools and methods in libraries and archives through professional education efforts. The project team will accomplish the following:

- **produce and disseminate** a publicly-accessible set of learning objects to be used by education providers in providing hands-on digital forensics education;
- **investigate and report** on institutional factors to facilitate, hinder and shape adoption of digital forensics educational offerings;

---

<sup>2</sup> For a more detailed summary of the following activities, see Lee et al, 2013, p.14-15.

<sup>3</sup> The earliest adopter in our Advisory Board is Patricia Galloway at the University of Texas, who offers courses at the iSchool in which students work in teams to address challenges associated with real-world digital collections. Beginning in 2005, she began setting the groundwork for a specialized Digital Archaeology Lab, which was established in 2009. Such units and adopters are rare, however, and most of the

<sup>4</sup> <http://www2.archivists.org/prof-education/das>

<sup>5</sup> <http://www.rarebookschool.org/courses/libraries/195/>

<sup>6</sup> <https://fido.cerch.kcl.ac.uk/digital-forensics-for-archivists-training-event/>



- **advance** a community of practice around digital forensics education, through partner collaboration, wider engagement, and exploration of community sustainability mechanisms; and
- **produce and disseminate** a white paper on strategies for offering digital forensics education for information professionals.

When we refer to “learning objects,” these are resources either produced by the project team or adapted (with permission and attribution) from the project partners that can be used in support of digital forensics education. These can include lecture materials, demonstrations of activities, discussion exercises, hands-on exercises, or assessment mechanisms. A major objective of BitCurator.edu is to produce materials that can be used in a variety of contexts (e.g. university courses, continuing education workshops, online settings, in-person settings). In order to promote reusability across different settings, we will be explicit about the technical resources required and prerequisite knowledge required on the part of the instructor and students. We recognize that reuse in an education setting will rarely mean wholesale copying. Instead, instructors must adapt resources to meet their pedagogical contexts, including the specific learning objectives they are pursuing (Friesen, 2004). In order to promote adaptability across different settings (graduate classrooms, online classrooms, workshops, etc.), we will be explicit about the learning objects, concepts and skills involved. For example, our default assumption will be that tasks be accomplished using freely available open-source software, but we will also expose users of the learning objects to proprietary tools that can accomplish the same tasks whenever possible.

### 2.1. Research Questions

The BitCurator.edu project will engage in research to investigate and support graduate-level LIS/preservation programs to implement educational offerings that make use of digital forensics tools and methods. This project will address two primary research questions:

**RQ1: What are the primary institutional and technological factors that influence adoption of digital forensics tools and methods in LIS classes in different educational settings?** Through interviews, surveys, and solicitation of various forms of community input, we will report on factors and implications for educational implementation strategies.

**RQ2: What are the most viable mechanisms for sustaining collaboration among LIS programs on the adoption of digital forensics tools and methods?** In close collaboration with the project partners, we will investigate several sustainability mechanisms. This will include, but not be limited to, potential establishment of a special educational member category within the BitCurator Consortium.

### 2.2. Partner Institutions

The Advisory Board represents a diverse set of schools: Catholic University, Indiana University, New York University (Moving Image Archiving and Preservation Program), San Jose State University, University of Illinois, University of Maryland, University of Michigan, University of Texas, Wayne State University. The BitCurator.edu project team has strong professional relationships with numerous people at these schools, including those holding senior and administrative positions. We believe those numerous connections will be strong assets to the project. In selecting individuals to serve as representatives on the Advisory Board, we have placed a heavy emphasis on faculty who have already made efforts to implement related coursework in their respective schools and are likely to be directly involved in implementing the BitCurator.edu materials (see Table 1).

Table 1 - Schools and Representatives on Advisory Board

School	Representative on Advisory Board
Catholic University	Jane Zhang, Associate Professor
Indiana University	Devan Donaldson, Assistant Professor
New York University	Howard Besser, Professor, Associate Director of MIAP, and Senior Scientist for Digital Library Initiatives for NYU Library
San Jose State University	Sandra Hirsh, Professor and Director of the School of Information
University of Illinois	Rhiannon Bettivia, Postdoctoral Research Associate
University of Maryland	Ricky Punzalan, Assistant Professor at iSchool, Affiliate Assistant Professor in Anthropology, and Co-Director of Museum Scholarship and Material Culture Program
University of Michigan	Paul Conway, Associate Professor
University of Texas	Patricia Galloway, Professor
Wayne State University	Kimberly Schroeder, Lecturer

All of the partner institutions have already taken steps toward incorporating elements of digital forensics education in their course offerings. However, they are not a monolithic set. The partners differ along several dimensions: institution type (private vs. public); unit size (faculty and students); institutional placement (self-standing school vs. unit within a school/department); curriculum areas of focus; educational deployment model (all in-person, all online, or hybrid); default educational software installation/hosting configurations (run directly on a local machine, run locally within a virtual machine, hosted by an external provider); IT support arrangements (provided by instructors themselves, dedicated IT staff within unit, centralized IT staff within institution, outsourced IT support providers); and level of adoption of digital forensics within curriculum. This diversity will help to ensure that our answers to both RQ1 and RQ2 are likely to be applicable to a wider population of education providers.

In addition to the project Advisory Board, we will also draw on a Professional Experts Panel, who will provide input and guidance on all aspects of the project through periodic conference calls:

- Bob Beatty, President, The Lyndhurst Group
- Tom Claeson, Senior Consultant for Digital Preservation Services, LYRISIS
- Robert Horton, Assistant Director for Collections and Archives, Smithsonian Institution
- Sarah Koonts, State Archivist of North Carolina, Council of State Archivists (CoSA)
- Nancy McGovern, Director of Digital Preservation, MIT Libraries
- Bethany Nowviskie, Executive Director of the Digital Library Federation (DLF)
- Susan Perry, Executive Director, Southeastern Museum Conference (SEMC)

In addressing RQ1, the project team will work with members of the Advisory Board to test and implement several models for administering digital forensics education, ranging from use of dedicated classroom space, to relying on student’s own laptops, to deployment of software functions over the Web. The project team will share guidance documents from these activities, which will serve as resources for education efforts elsewhere. RQ2 will be addressed through engagement with the BCC, Advisory Board and PEP on sustainability models, including (but not limited to) development of a new educational membership category for the BCC.

### 2.3. Plan of Work

The BitCurator.edu project will involve various efforts to advance the goals of addressing RQ1 and RQ2 and producing recourses for use by educators. Year 1 will include data collection and needs assessment from advisory board members, and development of educational resources to test. Year 2 will include an in-person Advisory Board meeting and public symposium; collaboration with Advisory Board members on testing of educational resources; and development and deployment of two webinars targeting educators at other institutions. In year 2, in collaboration with CoSA, we will generate: 1) two webinars - one as an introduction to digital forensics for public records professions, and another on how to teach digital forensics tasks to others (train the trainer); and 2) a workshop (tentative venue: State Archives of NC) for records professionals to educate others about using digital forensics methods to advance electronic recordkeeping objectives. Year 3 will also include the concluding Advisory Board meeting (administered remotely), public symposium and generating of final project deliverables.

Table 2 provides a summary of project activities, based on a quarterly schedule. The table is designed to identify major milestones and transition points between tasks. It does not reflect several ongoing activities, such as monthly conference calls with the Professional Experts Panel and Advisory Board, as well as active engagement with interested educators at other institutions. Also not referenced are outreach, engagement and educational activities in the form of conference presentations, tutorials and workshops. Such activities will represent a substantial time commitment by the team throughout the course of the project. They will be driven by both user demand and the contingencies of conference submissions being accepted, so we will not know the specific dates for such activities until the project is already underway. Note that throughout Years 2 and 3, we will be working with the project partners to iteratively refine and enhance the learning objects that they will be testing in their own settings.

Table 2 - Quarterly Activity Overview for BitCurator.edu (colors indicating project years)

Jul 1 – Sept 30 2018	On-boarding of all project participants, including webinars for partners on BitCurator environment, BitCurator Access tools, BitCurator NLP tools, and wider ecosystem of free and proprietary tools; formative project evaluation
Oct 1 – Dec 31 2018	Interviews with Advisory Board members on existing educational offerings, institutional adoption factors and desired offerings; analyze data and report findings
Jan 1 – Mar 31 2019	Survey targeting all educational providers (beyond project partners), with instrument based on findings from Advisory Board interviews
Apr 1 – Jun 30 2019	Completion of first draft of learning objects; preparation for Advisory Board meeting and public symposium
Jul 1 – Sept 30 2019	Advisory Board meeting and public symposium; analysis and reporting of outcomes
Oct 1 – Dec 31 2019	Host webinar, “Introduction to Digital Forensics for Public Records Professionals”
Jan 1 – Mar 31 2020	Host webinar, “How to Teach Digital Forensics Tasks to Others”
Apr 1 – Jun 30 2020	Preparation for year 3 Advisory Board meeting and public events
Jul 1 – Sept 30 2020	Advisory Board meeting (administered remotely), public symposium in Chapel Hill, and in-person workshop for records professionals at State Archives of North Carolina
Oct 1 – Dec 31 2020	Elicit feedback from Advisory Board and PEP on final draft learning objects
Jan 1 – Mar 31 2021	Final public release of learning objects; elicit feedback from Advisory Board and PEP on whitepaper, “Digital Forensics Instruction for Information Professionals”
Apr 1 – Jun 30 2021	Final public release of whitepaper and all other project documentation



## 2.4. Financial, Personnel and Other Resources

The project budget requested is \$499,664. This includes a) travel and accommodations for one in-person Advisory Board meeting, b) salary and benefits for the project manager and community lead (at the Educopia Institute), c) a portion of the time of the PI (Christopher Lee) and one of the Co-PIs (Katherine Skinner), and d) travel expenses for outreach and engagement. The project will also benefit from contributions from the Advisory Board represents and Professional Experts Panel (PEP) as described above. Letters of commitment are included from all project partners.

The following are brief highlights of project personnel and areas of expertise:

**Christopher (Cal) Lee** (Professor, UNC SILS) will act as principal investigator (PI) and overall coordinator of the project. He will interact with members of the Advisory Board and Professional Experts Panel, as well as overseeing administrative and financial aspects and reports. Lee is the founder and lead researcher of the BitCurator environment. He has served as PI or Co-PI for numerous grant-funded projects focusing on digital forensics software development, digital curation education, and digital curation workflows.

**Dr. Katherine Skinner** (Executive Director, Educopia Institute and Adjunct Professor) will act as co-principal investigator (co-PI) and overall coordinator for the project. She will oversee the work of the project manager at the Educopia Institute, as well as serving as a vital point of contact between the project and the BitCurator Consortium. Skinner has served as PI for grants and contracts totaling more than \$2.5M, including a broad range of cross sector initiatives in digital preservation across libraries, archives, and museums. Skinner is a founder of the MetaArchive Cooperative and the BitCurator Consortium, two leading efforts in digital preservation internationally. She will ensure the project and its deliverables adhere to open access principles and community frameworks, that they are both built and sustained by a range of committed partners, and that all outputs circulate across the memory organization landscape.

## 2.5. Outreach and Engagement

The project team is committed to wide dissemination of project findings and outputs. This outreach and dissemination process has already begun through our team's project planning work. It will continue to expand and amplify at the start of the project, building throughout the grant period and beyond. The project involves a diverse set of stakeholders, including representatives from different fields and perspectives. It will produce research findings, documentation, training materials, and a white paper, all of which will be distributed widely.

All project outputs will be published with CC-BY or CC-BY-SA or GPL v3.0 licensing and disseminated as freely and widely as possible. Due to our personal relationships throughout the field, our outreach efforts will not be limited to listserv distribution, but will also travel across our (intentionally diverse) networks through a wide variety of webinars, presentations, social media announcements, press releases, media outreach, and invitations targeted to library and archives directors, digital curation and preservation practitioners, trainers, and students to use the research outputs. We will encourage the use of outputs by a variety of continuing education and professional development groups throughout the nation and beyond, through such groups as the Society of American Archivists (Digital Archives Specialist curriculum), Coalition to Advance Learning (with more than 20 meta-organization members) and Educopia's project groups and Affiliated Communities (including the Mapping the Landscapes and Nexus LAB efforts, each with more than 35 meta-organizations that offer training to library, archives, and museum practitioners), and with existing trainers in this area to ensure broad uptake and reuse. We will actively reach out through the BitCurator Consortium, which is administered through Educopia and on which Lee serves as a member of the Executive Council. The BitCurator Consortium will host materials on its web site, and CoSA will do so through its Program for Electronic Records Training, Tools, and Standards (PERTTS Portal). The PEP will also help by reaching out through their various professional networks.

## 2.6. Risks

The project team recognizes that any project with multiple partners and stakeholders faces an array of potential risks. Specific risks and mitigation strategies we have actively planned include the following:

1. **Produce findings that lack relevance beyond the project partners** - We have selected partners that differ along several dimensions (see Research Design above for an elaboration of these dimensions).
2. **Failure to fully implement educational offerings at partner institutions** - As a project focused on institutional factors in adoption and models for sustaining collaboration, core project objectives will be met when partners define and share information about workflows. A large, diverse set of partners reduces dependency on any single institution's educational offerings to meet project objectives.
3. **Insufficient local IT support** – The project team will provide significant support for partner deployment. The BitCurator environment is widely run within a virtual machine, which requires very little change to local system settings or configuration. The BitCurator user community provides substantial peer technical support through its public mailing list.
4. **Lack of tools sustainability** – The focal set of tools are incorporated in the BitCurator environment, which is open-source software and associated documentation that is freely available online. The BitCurator Consortium advances the sustainability of the software and its associated user community. Use of OSS and a modular design approach within BitCurator minimizes future dependency on any specific software. Partners will share information about alternative tools for performing specific functions, which will be incorporated into learning objects, publications and the white paper.

## 2.7. Evaluation

We will work with an established Evaluator experienced in both formative and summative evaluation. The Evaluator will work with the project team at project kick-off to evaluate the project plan and start up activities against the outcomes and deliverables expected from the project. During this formative evaluation, the Evaluator will make recommendations for adjustments that will strengthen our approach and ensure the success of our work. We will maintain connection with the evaluator, conducting at least quarterly check ins on progress toward all project goals. Using summative evaluation frameworks and methods (e.g., interviews, surveys, logic models and outcomes-based evaluation), the Evaluator will conduct a rigorous evaluation of the project against its stated goals and outcomes. The resulting evaluation will be included in our final report.

## 2.8. Sustainability

It is critical that the project deliverables 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 will promote all review and final versions of project deliverables. As indicated above, all project information will be released with CC-BY or GPL v3.0 licensing to promote use and reuse with appropriate citation.

Given their demonstrated ongoing interest in advancing digital forensics education, our project partners and project team are well positioned to maintain and continue to build on these documents and tools in the course of future projects as well, as evidenced by the sustained efforts in previous years that this proposal builds upon. However, we find sustainability to be so essential to this endeavor that our research design includes RQ2, which is specifically focused on sustaining collaboration among LIS programs. We believe that one of the most promising outgrowths of these efforts is the BitCurator Consortium (BCC).

Since its formation in 2014, the BCC has served as a focal point and catalyst for professional engagement around the application of digital forensics in library and archives settings. It now has 27 member institutions from four different countries on three continents. The BCC has run a popular User Forum for the past three

years. As described in Section 1.1 above, online forums associated with the software are also very popular and active. BCC members also have access to education webinars and a dedicated member help desk. The BCC has just completed a detailed strategic planning process, with sustainability being a core focus of these discussions.

### 3. Diversity Plan

Two fundamental issues of diversity in collecting institutions are what they collect and how they provide access to what they have. The cultural record available through libraries and archives has systematic biases, which privilege certain populations and interests, while silencing or misappropriating others. Over the past several decades, there have been many efforts to acknowledge and potentially counteract some of these silences and distortions (Keough, 2002). The adoption of new technologies has the potential to both advance and hinder diversity objectives of libraries and archives in various ways (Daniel, 2010; Sadler and Bourg, 2015).

The incorporation of digital forensics tools and methods into digital curation workflows can dramatically shape both what information ends up in born-digital collections and how that information is exposed and presented to the world. To address these issues, two of the project's learning objects will be specifically focused on the diversity implications of applying digital forensics methods to digital collections. These materials will build off upon the previous work of the PI on this topic (Lee, 2010; Lee and Woods, 2013). It is our hope that development and dissemination of these learning objects should (1) raise awareness of diversity issues among instructors and students, and (2) attract a more diverse set of professionals to pursuing the work of applying digital forensics to collections when they see how it can address issues that resonate with them personally.

### 4. Broad Impact

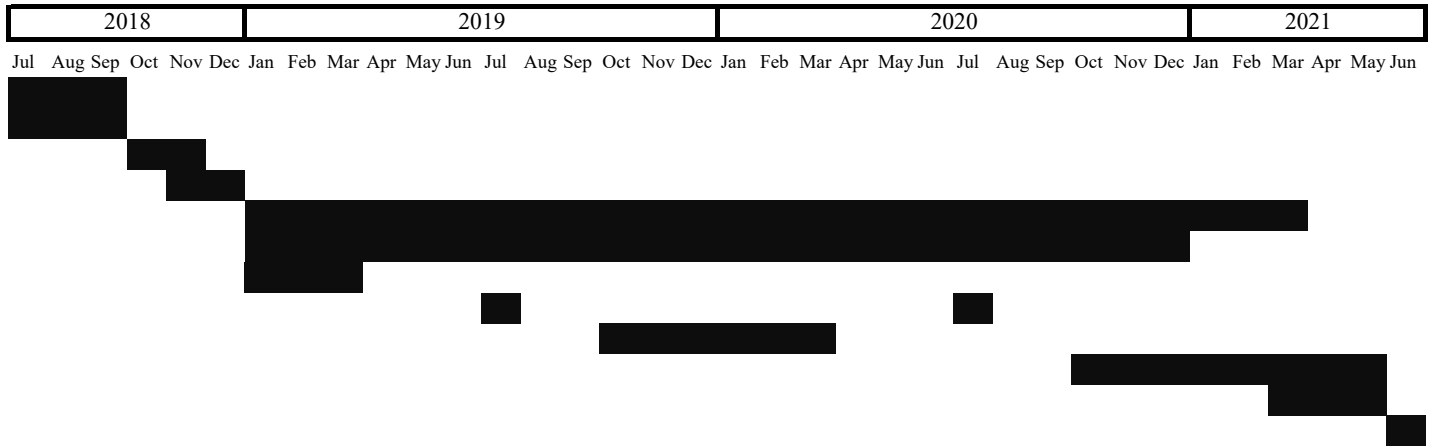
This project will dramatically advance the **National Digital Platform** by 1) producing generalizable findings about factors influencing various educational implementation approaches and strategies for addressing them, 2) building the capacity of educational programs across the country to better prepare their students to effectively manage, preserve and provide access to born-digital materials, and 3) providing exemplars and guidance on experiential learning opportunities that can serve as models for other institutions.

More specifically, this project will impact professional education (and, in turn, digital curation professional practices) by increasing our understanding of how institutions of different sizes and types choose to integrate digital forensics tools and methods into their educational offerings. Findings will support a broad range of institutions and individuals responsible for professional education. The knowledge gained from working with multiple institutions of different types and sizes will broaden understanding of educational approaches and priorities, and how those impact instruction in digital forensics and the curation of born-digital materials more broadly. We expect the empirical findings about institutional needs, as well as shared learning objects and synthetic white paper, to contribute substantially to digital curation research and practice. We expect the project to achieve the following measurable outcomes:

1. Strengthen information sharing across partners. **Evidence:** documentation of adaptation and/or reuse of learning materials between institutions.
2. Testing of learning objects by partners. **Evidence:** documentation of each partner's syllabi and other course materials
3. Adoption of learning objects by institutions beyond the partners. **Evidence:** examples elicited through professional forums and events.
4. Increase professionals' knowledge of and fluency in digital forensics instruction. **Evidence:** At least 40 trainees (webinars, workshop) self-report increase in knowledge and understanding.

**Activity**

- On-boarding and introductory training
- Formative project evaluation
- Advisory Board interviews
- Analyze, report on AB interviews
- Iteratively develop learning objects
- Partners test learning objects
- Survey of educational providers
- AB meeting and public symposium
- Create and run targeted webinars
- Develop white paper
- Summative evaluation
- Wrap up and final reporting



# DIGITAL PRODUCT FORM

## Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

## Instructions

- Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

If you ARE creating digital products, you must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

## Part I: Intellectual Property Rights and Permissions

**A.1** What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

*All content, resources, and assets produced by the project and intended for distribution to the community will carry the Creative Commons BY-SA 4.0 license. These licenses were selected to maximize community uptake and potential for remix and reuse in future products.*

**A.2** What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

*No conditions will be imposed on access or use of the primary products of the project beyond those specified by the licenses CC-BY-SA 4.0 and GPL v3.0.*

**A.3** If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

*This project will explore tools and methods that may apply to private and sensitive datasets, but will not collect or otherwise distribute those datasets. Any public demonstrations of the techniques during the project will use only synthetic datasets.*

## Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

### A. Creating or Collecting New Digital Content, Resources, or Assets

**A.1** Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

OMB Control #: 3137-0092, Expiration Date: 7/31/2018

IMLS-CLR-F-0032



*This project will product learning objects (lecture materials, demonstrations of activities, discussion exercises, hands-on exercises, or assessment mechanisms), a project white paper, and various presentations, workshop materials and published articles about the project.*

**A.2** List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

*The project team will use standard hardware and software available to them through their institutions. All software will either be open-source or provided through licenses already available.*

**A.3** List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

*Project reports will be distributed in PDF format and comply (at minimum) with PDF/A-1a.*

## **B. Workflow and Asset Maintenance/Preservation**

**B.1** Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

*General project tasks and workflow will be managed using Asana (<https://asana.com/>). The development team at UNC SILS has successfully used this platform to manage multiple prior grant-funded projects.*

**B.2** Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

*During the award period, documentation assets will be maintained in both private and public cloud services. These will include Google Docs for contributions shared with project partners and Microsoft OneDrive for those assets internal to the development team at UNC SILS; OneDrive is currently the default cloud institutional asset management system for UNC SILS. All project documentation subject to public release will also be distributed in a format minimally compliant with PDF/A-1a via the BitCurator Consortium.*

## **C. Metadata**

**C.1** Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

*This project will not produce any technical, descriptive, administrative, or preservation metadata related to collections. It will provide guidance on how to administer and transform these types of metadata when working with born-digital materials.*

**C.2** Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

*This project will not create or maintain structured metadata aside from that metadata associated with learning objects and project documents, which will be stored alongside the documents on the project and BitCurator Consortium web sites.*

**C.3** Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

*This project will not use any specialized APIs. Those wishing to harvest metadata can use standard REST mechanisms.*

## **D. Access and Use**

OMB Control #: 3137-0092, Expiration Date: 7/31/2018

IMLS-CLR-F-0032

**D.1** Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

*Project documentation for public release will be openly available online. To avoid access bottlenecks or single points of failure, the project team will ensure that this documentation is made available (during the funded project period where required and in its entirety on project completion) via existing portals (including <https://www.bitcurator.net> and <https://www.bitcuratorconsortium.org/>).*

**D.2** Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

*URLs are provided below. Note that this is a representative (not complete) accounting of the projects associated with the primary contributors named in this grant:*

<https://educopia.org>  
<http://metaarchive.org/>  
<https://librarypublishing.org/>  
<https://www.bitcuratorconsortium.org/>  
<https://www.bitcurator.net/>  
<https://wiki.bitcurator.net/>  
<http://bitcurator.github.io/>

### **Part III. Projects Developing Software**

#### **A. General Information**

**A.1** Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

*This project will not produce software.*

**A.2** List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

*Not applicable.*

#### **B. Technical Information**

**B.1** List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

*Not applicable.*

**B.2** Describe how the software you intend to create will extend or interoperate with relevant existing software.

*Not applicable.*

**B.3** Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

*Not applicable.*

**B.4** Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

*Not applicable.*

**B.5** Provide the name(s) and URL(s) for examples of any previous software your organization has created.

*A complete list of projects developed by the BitCurator team at UNC SILS can be found at:  
<https://bitcurator.github.io>*

### **C. Access and Use**

**C.1** We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

*Not applicable.*

**C.2** Describe how you will make the software and source code available to the public and/or its intended users.

*Not applicable.*

**C.3** Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL:

*Not applicable.*

## **Part IV: Projects Creating Datasets**

**A.1** Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

*We will generate data from interviews of partners and surveys of (on-partner) education providers.*

**A.2** Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

We will seek approval from the IRB at the University of North Carolina at Chapel Hill. Applications are submitted through IRBIS: <http://research.unc.edu/oris/apps/irbis/>

**A.3** Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

*The data we collect will not be sensitive, and we will indicated in our informed consent documentation that anonymity of respondents will not be guaranteed.*

**A.4** If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

OMB Control #: 3137-0092, Expiration Date: 7/31/2018

IMLS-CLR-F-0032

*Not applicable.*

**A.5** What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

*We will use a written protocol for interviews. Surveys will be conducted using Qualtrics (<https://software.unc.edu/qualtrics/>).*

**A.6** What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

*We will retain the data collection instruments (interview and survey) and retain them with the data. We will deposit the data in the Carolina Digital Repository (<https://cdr.lib.unc.edu/>). Permanent association will be ensuring through the identification system of the repository.*

**A.7** What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

*Data will be deposited into the Carolina Digital Repository.*

**A.8** Identify where you will deposit the dataset(s):

*Name of repository: Carolina Digital Repository*

*URL: <https://cdr.lib.unc.edu>*

**A.9** When and how frequently will you review this data management plan? How will the implementation be monitored?

*The data management plan will be reviewed upon the initiation of the project and then at the beginning of each of the following two project years. Implementation will be ensured by the PI at UNC, and it will also be addressed in the work of our external evaluator.*