Sustaining digital community collections:
Understanding the Impact of Development Workflows Across Contexts

Digital collections created by humanities scholars and community archives collectively constitute an extensive, scattered body of evidence of cultures and histories that tend to be underrepresented in cultural institutions. Dwelling outside of memory institutions, these collections confront major barriers to sustainability, compromising the completeness and equity of our collective digital memory. "Sustaining Digital Community Collections" is a three-year Early Career Development research project to investigate collaborative workflows of collection development and maintenance, and to identify roles that cultural heritage institutions may play to help realize community-determined, community-led strategies for sustaining digital collections. This project will develop and evaluate: (1) Models of emergent development and maintenance workflows for digital community collections; (2) Roles for libraries and archives in those workflows; and (3) A toolkit to help stakeholder communities and memory institutions implement community-centered strategies for collection sustainability.

1. Statement of Broad Need. The sustainability of digital cultural collections is an urgent problem that transcends collecting institutions and independent communities of research and memory. The need for sustainable infrastructures built on values of openness, equity, and partnership has been acknowledged as a grand challenge for knowledge ecosystems of the future (Altman and Bourg, 2018). To advance the sustainability of community-based collections, this project responds to a need to investigate (a) sustainability challenges across collections in different contexts; (b) novel roles for memory institutions in community-centered sustainability; and (c) the impact of development and maintenance workflows on collection sustainability.

(a) Sustainability across collections in different contexts. A growing diversity of digital cultural collections exist outside of mainstream memory institutions, centered within the research and public communities that create them. Libraries and archives cannot and should not comprehensively collect and sustain the growing mass of digital community collections—particularly where communities wish to retain control over the ongoing care and use of collections. Both community archives and digital humanities collections are *community-centered*, in the sense that communities of shared interest develop these collections for their own purposes (Flinn et al., 2009; Palmer, 2004).

Community archives develop around shared identity, memory, and purpose. Communities develop digital archives to share experiences of localized histories and places; significant events; social justice and activist initiatives; an ethnicity, race, or religion; gender identities and sexual orientation; etc. (Welland and Cossham, 2019; Caswell, 2014a; Flinn et al., 2009). For example, the *Lakeland Digital Archive* is an effort of the Lakeland Community Heritage Project[1] to document a local, historic African American community by digitizing family records and gathering oral histories. At the same time, within and outside of the academy, individual humanities scholars and small, collaborative teams of technologists and researchers are gathering and maintaining their own collections of digital primary sources and data (Poole, 2017; Cooper and Rieger, 2018; Palmer, 2004). For example, *Enslaved: Peoples of the Historic Slave Trade*[2] is a linked open data platform for the study of the historic slave

[1] https://lakelandchp.com/
[2] http://enslaved.org/

trade, and the *Open Islamicate Texts Initiative*[3] is creating a corpus for the study of texts from the premodern Islamicate world. Digital humanities collections serve the purposes of constituents of these teams and wider communities within the same research domain or discipline (Maron and Pickle, 2014; Benardou et al., 2013); and broad public impact is a core value of digital humanities practice (Spiro, 2012). Both digital community archives and digital humanities collections are developed to fill gaps in representation in mainstream institutions. Their value derives not only from the original evidence they gather and the rich contextual information they build on top of that evidence, but from their community-centeredness: their unique capacity to represent the voices or serve the distinctive needs of the communities that create and use them (Flinn, 2007; Fenlon, 2017).

Sustainability is particularly problematic for community-centered collections. Their maintenance often depends on the work of small, under-resourced communities. They tend to be funded sporadically, if at all, and are often built on fragile technical infrastructures. Without systemic support, they tend to decline as undergirding technologies or the communities themselves shift. While these collections may fall within the scope of the preservation mission of libraries and archives, they resist exclusively institutional models of sustainment, which often rely on outright transfer of a collection to a memory institution. Many community archives value autonomy. They wish to retain control, to align access to and representation of their materials with community values (Caswell et al., 2017). Digital humanities projects, on the other hand, generate value by serving as hubs for active and ongoing collaboration and communication within research communities (Fenlon, 2020). Their resources are sustained not by transfer from communities to preservation institutions, but through continued life and development. More pragmatically, most memory institutions lack the capacity to comprehensively collect and preserve digital community collections. Identifying community-centered sustainability strategies emerges as an urgent priority for the completeness and equity of the digital cultural record.

To sustain a distributed cultural record across a diversity of communities and institutions, various stakeholders in cultural heritage—including public and academic libraries and archives, historical societies, research communities in the humanities, and community archives—need a broader, foundational understanding of the challenges confronting collections and communities across different contexts. Communities that create digital cultural collections come in all sizes, organized around different nuclei of interest. They grow under a vast range of conditions—of resourcing, governance, labor, scale, support, and within different organizational, geographic, sociopolitical contexts. All of these contextual factors shape those collections and impact their sustainability. While there is an emergent body of work on community archives sustainability (Newman, 2011; Froese-Stoddard, 2014; Du Laney, 2019) and parallel, emergent work on sustaining digital humanities scholarship (Smithies et al., 2019; Langmead et al., 2018; Edmond and Morselli, 2020), there is a dearth of empirical and conceptual work that looks across contexts to identify transcendent challenges and broadly useful strategies rooted within communities.

(b) Roles for memory institutions in community-centered sustainability. In practice, the main role that libraries and archives play in stewarding community collections entails the transfer of ownership from communities to institutions. For this reason, research and practice on library roles in sustaining digital collections tend to focus on institutional problems of organizational resilience, project management, and economic viability (Eschenfelder et al., 2016); or on defining policies for the collection and management of digital objects, including service-level agreements and procedures for

[3] https://iti-corpus.github.io/

deposit and migration (Smithies et al., 2019; Oltmanns et al., 2019; Madsen and Hurst, 2018; Goddard and Walde, 2017). These contributions offer significant advancements in institutional practice, but because they are focused on individual institutions rather than communities, they cannot address a system-wide challenge confronting both communities without recourse to a sustainable institutional backer, and independent communities that wish to keep their collections autonomous.

Cultural institutions are moving toward promising new models of partnership with communities and shared stewardship of collections (Flinn, 2011; Cook, 2013; Caswell, 2014b; O'Neal, 2015). For example, the Smithsonian Center for Folklife and Cultural Heritage recently announced a *Shared Stewardship of Collections Policy,* which describes models of co-curation, long-term loan, and digital return as alternatives to traditional collection-transfer (Smithsonian, 2019). Institutions are developing policies for fostering lasting partnership with communities, and there are scattered efforts to provision community sustainment through the reallocation of collection development funds, professional consultation, and spaces for building trust and equitable partnership. Institutions are creating resources such as workshops and best practices to support community curation work, such as the Socio-Technical Sustainability Roadmap.[4] These exciting developments suggest the need for and timeliness of research on the range of practices and roles available to institutions to support community efforts in novel ways, beyond stewarding their collections. In particular, existing work on institutional interventions in community sustainability stems predominantly from within preservation institutions and professional practice. Libraries and communities alike need a foundational understanding of potential roles in sustainability based on research that is situated in and focused on communities and collections in different contexts.

(c) Impact of development and maintenance workflows. It is widely acknowledged that decisions made in the development stages of projects—about the organization of records, descriptive practices, which file formats or technical standards to adopt, etc.—have an inordinate impact on the ultimate sustainability and preservation of collections. But how the *processes* or *workflows* of development and maintenance affect sustainability is less well understood. A digital collection is constructed and maintained through collaborative, manual and technical processes, including specific procedures for digitization, ingestion, description, curation, migration, etc. These processes are often invisible, undocumented, and complex. Preliminary research found that keeping a collection or resource viable entails more than stewarding the collection's digital artifacts and technical infrastructure; it also entails managing the amorphous social infrastructure that serves to maintain the collection (Fenlon, 2020). This finding resonates with a growing recognition of the sociotechnical aspects of sustainability and maintenance work (Edmond & Morselli, 2020; Langmead et al., 2018; Madsen and Hurst, 2018).

A long strand of research in library and information science has sought to create abstract models of researchers' information processes and activities. These models aim to inform development of cultural collections and humanities research infrastructure to support researchers' existing practices across a variety of domains (Ellis, 1993; Unsworth, 2000; Palmer et al., 2009; Blanke and Hedges, 2013; Vilar, 2015; Benardou et al., 2013; Gradmann et al., 2015). Research on information practices and information seeking has sought to distill a set of primitives, or basic information activities and processes that are common across researchers and disciplines, such as "browsing", "chaining", "note-taking", "networking", "gathering", etc. (Ellis, 1993; Unsworth, 2000; Palmer et al., 2009; Blanke and Hedges,

[4] https://sites.haa.pitt.edu/sustainabilityroadmap/

2013; Vilar, 2015). Building on this strand of work, recent efforts including the Scholarly Domain Model (Gradmann et al., 2015) and the Scholarly Research Activity Model (Benardou et al., 2010; Benardou et al., 2013) offer more general accounts of scholarly practices, which represent specific activities as components within complex systems of actors, social contexts, domains, methods, goals, tools, etc. Individual researchers combine these activities into *workflows* or overarching processes (Antonijevic and Cahoy, 2018). The picture becomes more complicated when the workflows are collaborative.

Formally describing, characterizing, or mapping these workflows can identify problem points, facilitate preservation work and provenance description, and lead to the development of more effective information systems. Recent work in this vein has explored the benefits of documenting workflows of data migration to guide preservation efforts in natural history museums (Thomer, 2020), and illuminated the increasing prevalence and importance of shareable research workflows in humanities scholarship (Schonfeld and Waters, 2018; Liu et al., 2017). While there is a substantial body of work on collection-development workflows in libraries and archives, largely focused on digitization and processing, workflows have received less attention in collaborative and community-centered collections. How workflows affect sustainability in different contexts is understudied, and in particular, there is a need for research bridging workflows of development and those of maintenance.

Communities developing digital collections, along with libraries and archives seeking to support those communities, would benefit from general models of common, community-centered collection workflows in two ways. First, mapping out workflows helps expose processes (technical and manual) that are resource-intensive, vulnerable to breakage, invisible and undocumented, or otherwise problematic from a sustainability perspective. On the same token, workflow-mapping can reveal aspects of collection development and maintenance that might benefit from library intervention or support. Second, community archives and digital humanities projects would benefit from the availability of models of common practices, which could be adapted to their own projects (Caswell et al., 2017). This project seeks to build on existing guidance and best practices for community archives and digital humanities projects, such as the Socio-Technical Sustainability Roadmap[5] and the Archives-in-a-Backpack project.[6] The proposed project aims to characterize models of development and maintenance workflows for digital community collections, and articulate roles for libraries and archives in those workflows, with the end goal of increasing capacity in both memory institutions and communities for community-centered sustainment.

2. Project design. This three-year project seeks to answer the following questions: (1) How can we identify, characterize, and represent sociotechnical workflows for the development and ongoing maintenance of digital collections, including digital humanities collections and community archives? (2) How do these workflows affect the sustainability of collections in different contexts? (3) What are roles and best practices for libraries, archives, and stakeholder communities in supporting community sustainability? This project will conduct a multi-phase study of digital collections and their communities. To this end, the project requests funding to support one full-time graduate research assistant (including tuition and salary/wages), two hourly student research assistants, travel for site visits and conference dissemination, materials and supplies, and PI salary and fringe benefits.

[5] https://sites.haa.pitt.edu/sustainabilityroadmap/

[6] https://library.unc.edu/wilson/shc/community-driven-archives/archivist-in-a-backpack/

This project is designed as a comparative, multi-case study of three central cases, mentioned above: (a) the Enslaved project; (b) the Open Islamicate Texts Initiative; and (c) the Lakeland Digital Archive. The goal of case study research is to develop a deep understanding of a phenomenon set within its real-world context (Yin, 2018), which is well suited to the highly context-dependent problem of sustainability. Each case selected for this study encompasses a community actively developing a digital collection about people, histories, or textual traditions that are underrepresented in mainstream institutions. They collect original, unique forms of digital evidence, ranging from photographs and maps to linked data. Each community bears different relationships to relevant libraries and archives. Because these collections are undergoing active development, they afford opportunities to closely examine development workflows and decisions that impact sustainability, and to directly engage relevant stakeholders in conversations and decisions about the roles of libraries, archives, and their own communities in the future of each collection. The PI has an established research relationship with each community, and all have given assurances of ongoing support, access, and partnership in this work.

These three central cases, focused on community-centered collections, will be augmented with (a) interviews with representatives of libraries and archives that are or could be connected to each case, and (b) a set of additional "mini-cases" of long-running community-centered collections, to complement the in-development cases with retrospective analysis of demonstrated sustainability strategies. Through cross-case comparison and analysis, this project will develop and evaluate:

(1) Models of development and maintenance workflows for digital humanities research collections and digital community archives, which can be reused in other digital collection projects, and which illuminate potential sustainability vulnerabilities at the stage of development;
(2) Roles for libraries and archives in those workflows; and
(3) An open toolkit to help stakeholder communities navigate their own sustainability challenges and available strategies, and to help memory institutions understand their potential roles in supporting community sustainability.

Table 1. Overview of project design

Year	Phase	Outcomes
1	Develop case studies of community collections and their development workflows	● 3 complete case studies (including interviews, documentary evidence, and participant-observation) focused on documenting workflows ● Preliminary set of development workflows identifying points of vulnerability
2	Build cross-case comparisons with established community collections and memory institutions	● Interviews with relevant memory institutions ● 5-10 additional "mini-cases" of long-running community-centered collections ● Cross-case analysis to produce a refined and expanded set of workflows
3	Develop and disseminate sustainability toolkit for	● Case partner evaluation of workflows ● Final set of model workflows for collection

	communities and memory institutions	development and maintenance ● Open toolkit for stakeholder communities and memory institutions

Cases

The three cases—the Lakeland Digital Archive, the Enslaved project, Open Islamicate Texts Initiative—were selected to facilitate rich comparison and contrast among collections that exhibit significant differences in their communities, context, scope and scale, technical architecture, and funding histories. They exemplify projects that make unique contributions to the equity and inclusion of our cultural records, through the collection of evidence about underrepresented people and histories. The selected cases also leverage existing research relationships centered at the University of Maryland; local connections have afforded the PI unique access and the opportunity to build upon existing relationships of trust and collaboration. The PI has established contact and partnership with all three cases and engaged them in preliminary data collection beginning in 2019. Through an initial round of interviews with a subset of representatives from each project, the PI has established a foundational understanding of each project's team, goals, operations, and outlook. The data collection activities described below, designed to develop the cases toward the focus on workflows and roles for libraries and archives, will build upon and incorporate this preliminary data collection. IRB approval has been obtained for all three cases and is included among the supporting documents.

Case 1: Lakeland Digital Archive: Lakeland is a 130-year-old African American community adjacent to the University of Maryland in College Park. For the past decade, its residents have worked to collect and preserve their cultural heritage as part of an overarching story of African American life in the long 20th century. Building on an existing community-university research relationship, the Lakeland Community Heritage Project[7] recently embarked on a partnership with the Maryland Institute for Technology in the Humanities[8] (MITH) to prototype a digital community archive. With a dedicated team of approximately one dozen people (approximately half from the Lakeland community and half from MITH), the project has assembled a large and growing collection of photographs, papers, audio and video, maps, etc., through digitization workshops and oral history initiatives. The infrastructure for this digital archive is under development, and will apply minimal computing methods such as static sites and simple scripting[9] to reduce the burden of maintenance and foster community-centered sustainability. This case affords the opportunity to engage not only the team developing the project (many of whom are Lakelanders), but also the wider community represented in the digital archive, in interviews and dialog about the collection's future.

Case 2: Enslaved project: *Enslaved: Peoples of the Historic Slave Trade*[10] is a distributed, collaborative effort based at Michigan State University's MATRIX[11] center, to aggregate historical data about enslaved people into a large-scale linked open data hub. The hub aims to integrate heterogeneous and

[7] https://lakelandchp.com/
[8] https://mith.umd.edu/
[9] See Gil and Risam's (2019) definition of minimal computing in this call for proposals: http://www.digitalhumanities.org/dhq/submissions/cfps.html
[10] http://enslaved.org/
[11] https://msustatewide.msu.edu/Programs/Details/1219

scattered—but potentially interconnected—datasets produced by historians, to facilitate identification and disambiguation of people involved in the slave trade and support historical and slave studies scholarship and public family-history research. The hub is in development, scheduled for public release in 2020. Its development team is small, but the project is engaging a wider set of partners and building a community of stakeholders among scholars and the public. This case affords the opportunity to study sustainability challenges related to linked data aggregations and distributed partnerships.

Case 3: Open Islamicate Texts Initiative: OpenITI[12] is a multi-institutional effort established in 2016 to construct a machine-actionable corpus of premodern Islamicate texts. Led by researchers at the Aga Khan University, Universität Leipzig, and the Roshan Institute for Persian Studies at the University of Maryland, the goal of this project is to develop textual infrastructure—including a corpus and relevant tools—in Persian and Arabic, to facilitate new forms of computational and macro-scale text analysis and digital scholarship. To this end, OpenITI has embarked on a concerted interdisciplinary community-building effort, engaging scholars in the fields of Islamic, Persian, and Arabic Studies, and in digital humanities. OpenITI's ultimate goal is to comprehensively digitize the textual traditions of the premodern Islamicate world.

To develop each case, the project team will engage case partners in semi-structured interviews and participant-observation, and analyze documentary evidence of each community and collection, including technical documentation, strategic and planning documents, meeting notes, and the digital collections themselves. The project team will visit each project in situ through site visits, with the goals of forging stronger relationships with case partners to increase information sharing, increasing the likelihood of serendipitous discovery, and gaining a more holistic view of each project than can be gained through more limited modes of access. In addition, case partners will be engaged in iterative evaluation and refinement of the workflows and toolkit that this project aims to develop, as described in year-by-year activities below. Case participation is designed to mutually benefit the case partners and this research project. Their participation will ensure the representativeness and impact of the toolkit across a broad range of collections and contexts. In turn, participation will directly help case partners document their own workflows, articulate their sustainability challenges, and implement community-centered sustainability strategies.

Year 1. Develop case studies of community collection workflows in active development. The project team will develop case studies of each collection in development, with a focus on characterizing sociotechnical, collaborative workflows of development and maintenance, including for the project as a whole and for specific processes such as metadata-creation, data migration and transformation, design, collection development, etc. Case studies will be developed through (a) analysis of documentary evidence, including internal documentation and each collection itself; (b) site visits to conduct participant-observation in development meetings (one to the remote site of the *Enslaved* project, and the other two local site visits); and (c) interviews of the dedicated teams developing the collections as well as relevant stakeholders, including the broader communities represented by or invested in each project. The project team will draft initial models of workflows, identifying points of vulnerability for sustainability. Toward the end of year 1, the project team will begin disseminating interim project developments and gathering feedback through relevant research and practitioner conferences.

[12] https://iti-corpus.github.io/

Year 2. Build cross-case comparisons with established community collections and memory institutions. The project team will augment the three central cases with interviews with representatives of relevant university libraries and archives that have potential stakes in each collection. In addition, through semi-structured interviews, focus groups, and analysis of documentary evidence, the project will build 5-10 additional small-scale cases of established or long-running digital humanities collections and digital community archives. These "mini-cases" will target collections and communities with a variety of sustainability models, and different kinds and degrees of relationships with memory institutions. Cross-case analysis will be conducted across the full set of cases and supplementary evidence, to refine the preliminary workflows and identify roles for libraries and archives in sustainability strategies. The project team will continue disseminating interim project developments to the case partners and through relevant research and practitioner conferences. If representatives of the collections selected as mini-cases are present at conferences, conference travel will support in-person interviews and focus groups with these representatives. Otherwise, data collection for the mini-cases will be conducted remotely.

Year 3. Evaluate workflows and develop toolkit to support communities and libraries. The team will formally evaluate workflows in partnership with the case study communities through an iterative feedback process. The project team will then create and publish an open toolkit for stakeholder communities and memory institutions to map equitable workflows and context-responsive, community-centered strategies for sustainable collection development and maintenance. The toolkit aims to integrate all project outcomes, including the cross-case analysis, feedback and evaluation from case partners, and the workflows. While the precise form of the toolkit will emerge from the research, it will include the following general components: (a) a summary of sustainability concerns confronting community-centered collections in different contexts, with guidance on how other communities have confronted this challenge; (b) models of workflows of different aspects of digital collection development in different contexts (including community archives and digital humanities collections), with points of vulnerability identified, and points where library partnership may contribute to community sustainability; (c) guidance for other communities on mapping their own workflows to support community decision-making; (d) guidance on context-driven strategies for community-centered sustainability, as identified in the literature and in the course of research; (e) guidance on roles for libraries in creating equitable library/community partnerships, or in supporting community sustainment of digital collections; and (f) guidance to other relevant resources to support community-centered sustainment. The toolkit will be made available as an openly accessible website and deposited into the University of Maryland's institutional repository for preservation and ongoing. The toolkit and all project outcomes will be broadly disseminated through the project website and social media, journal publication, and professional and research conferences oriented toward library and archives practitioners and communities developing digital collections.

Personnel

The graduate research assistant (anticipated to be a library and information science PhD student at the University of Maryland) will provide dedicated support to all aspects of the research while gaining experience and expertise in community-centered digital collections, research methods, project management, and sustainability. The graduate assistant will contribute to participant recruitment, data collection and analysis, and dissemination, and lead an ongoing review of relevant developments in theory and practice in the domains of library and information science, community archiving, science and technology studies, computer-supported cooperative work, community endurance, and related fields. Over the course of the project, the two hourly student positions are intended to include both master's

students in library and information science and undergraduate students from the unprecedented Digital Curation Specialization in the University's Information Science undergraduate major. This specialization is designed to recruit students from diverse backgrounds into the master's program and careers in cultural institutions. These students will gain hands-on experience with many facets of digital curation and collaborate with the graduate assistant and PI on data collection and analysis, workflow visualization, and website development to support toolkit dissemination.

The PI will lead case study development, coordinate with case partners, guide students in data collection and analysis, lead the development of workflows based on data analysis, and lead the toolkit development and dissemination of project outcomes. Dr. Fenlon has experience in the design and completion of qualitative studies of information practices, digital scholarship, and library, archive, and museum collections. She has published in *Library Trends* and the *Journal of Scholarly Publishing,* and in conference proceedings including the ACM/IEEE Joint Conference on Digital Libraries and the Association for Information Science and Technology Annual Meeting. This project builds on her dissertation, completed in 2017, and constitutes the centerpiece of her research program, which aims to advance the capacity of communities and knowledge organizations to create and sustain meaningful, inclusive, impactful digital cultural collections that, in turn, support the advancement of knowledge and the endurance of communities. The budget justification provides detail on all personnel responsibilities.

Risks to project completion and mitigating factors

The project timeline could be affected by disruption to one or more of the case studies due to a variety of factors, such as case partners experiencing unexpected time or resource constraints that limit their participation and responsiveness, or a change in key personnel within the case study projects. This factor has been mitigated through a careful recruitment process that clarifies the project timeline and makes transparent the expectations of and for project partners, including the potential benefits to them. Backup cases are planned, should a case partner need to leave the study. An additional risk to the completion of this research has emerged in the form of the current global pandemic, the trajectory of which remains uncertain. To mitigate this risk, the PI has established processes for working with research assistants exclusively online, including weekly or more frequent videoconferences and the use of shared collaborative workspaces. The PI also has established processes for conducting interviews online. Conference travel, community interviews, and site visits can be delayed until the pandemic has passed, or replaced with more sustained, virtual forms of engagement.

3. Diversity plan. Because the driving impetus for this project is to sustain collections of underrepresented cultural records within a growing diversity of communities and institutions, recruiting researchers and participants from underrepresented populations in information professions is an express priority of this project. The University of Maryland College of Information Studies (iSchool) has established infrastructure to promote diversity and inclusion, having pioneered a master's level specialization in diversity and inclusion, and hosting an annual Conference on Inclusion and Diversity in Library and Information Science (CIDLIS), now in its 11th year.[13] This project is designed to benefit from the existing infrastructure for diversity and inclusion at the iSchool. In particular, the project's staffing model aims to increase opportunities for involving multiple students from across iSchool programs, including our unprecedented digital curation specialization for undergraduates, which has a highly diverse student population, and which was developed to open pathways for diverse students into library and archive professions. The PI is committed to recruiting student researchers from

[13] https://cidlis.umd.edu/

underrepresented backgrounds and mentoring their development in the information professions. In addition, the project is designed to engage diverse case partners in a mutually beneficial model of research partnership, which actively increases the sustainability of their community collections, which in turn contribute to redressing gaps in representation in our cultural records. The research methods, particularly the incorporation of participant-observation, are designed to foster dialog and real collaboration among the research team and case partners, and to maximize participants' ability to define sustainability challenges and workflows in their own terms. Participants will have repeated opportunities to provide feedback on project outcomes at each stage.

4. Broad impact. This project aims to increase the capacity for a broad set of stakeholders in cultural heritage to sustain a greater diversity of collections within a greater diversity of communities and institutions. Stakeholders include public and academic libraries and archives; scholars and research communities in the humanities; and community archives. By contributing to the longevity of digital collections within communities, the project ultimately aims to expand public access to cultural and historical knowledge, sustain community memory, and support library-community partnerships. The main vector for impact in this project is the toolkit, designed for the dual purpose of serving communities planning for the sustainability of their own collections, and libraries seeking new modes of partnership around community-based collections. Communities will be able to use the toolkit to map and improve their development workflows toward increasing the sustainability of their collections, as well as navigate potential partnerships with memory institutions. Library and archives professionals and students will be able to use the toolkit to discover and adopt new roles in supporting community-centered sustainability strategies. To ensure the toolkit is broadly useful, the project is designed to produce a user-centered toolkit that has been refined by iterative rounds of testing and feedback from case partners and library and archives professionals. The toolkit will directly benefit case partners, which will help ensure its broader implementation. Once published, the toolkit will be widely shared through presentations at professional and academic conferences.

In addition to developing the toolkit, the project aims to advance academic knowledge in the domains of library and information science, scholarly communication, and digital humanities, by generating insights about digital collections, their communities, and sustainability that transcend institutional and disciplinary contexts. Dissemination efforts are planned to happen through each phase of the project, for the purposes of increasing impact and also garnering feedback to improve the resulting toolkit. Dissemination efforts will target conferences and journals oriented toward library and archives professionals and researchers, in addition to conferences oriented toward communities developing digital collections, including community archives and digital humanities conferences, to be identified in consultation with case partners. Target conferences include the Digital Library Federation Forum (DLF), Digital Preservation (DigiPres), the Society of American Archivists (SAA), the Archival Education and Research Institute (AERI) Association for Information Science and Technology Annual Meeting (ASIS&T), Digital Humanities (DH), and others. Access to the toolkit will be sustained beyond the project period via the University of Maryland's open-access institutional repository[14] and through the project website. Because the proposed project is central to the PI's broader research program, the PI is committed to refining and enhancing the toolkit after the end of the budget period, and through future research continuing to expand the impact of this work.

[14] https://drum.lib.umd.edu/

Schedule of Completion
September 1, 2020-August 31, 2023

		Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
	Activities												
Year 1	Initiate data collection for all three cases	■	■	■	■	■	■						
	Site visits to each site			■	■	■	■	■					
	Begin preliminary data analysis			■	■	■	■	■	■	■	■		
	Begin workflow mapping								■	■	■	■	■
	Dissemination							■	■	■	■	■	■
Year 2	Initiate mini-cases	■	■	■	■	■							
	Begin interviews with memory institutions					■	■	■	■	■			
	Begin evaluation of preliminary workflows with case partners	■	■	■									
	Conduct cross-case analysis				■	■	■	■	■	■	■	■	■
	Refine workflows based on feedback and additional data										■	■	■
	Dissemination	■	■	■	■	■	■	■	■	■	■	■	■
Year 3	Case partners evaluate final workflows	■	■	■									
	Toolkit creation		■	■	■	■							
	Case partners evaluate and adapt toolkit to their contexts						■	■	■	■			
	Toolkit refinement and publishing								■	■	■	■	■
	Dissemination	■	■	■	■	■	■	■	■	■	■	■	■

DIGITAL PRODUCT FORM

INTRODUCTION

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to digital products that are created using federal funds. This includes (1) digitized and born-digital content, resources, or assets; (2) software; and (3) research data (see below for more specific examples). Excluded are preliminary analyses, drafts of papers, plans for future research, peer-review assessments, and communications with colleagues.

The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public. 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

If you propose to create digital products in the course of your IMLS-funded project, you must first provide answers to the questions in **SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS.** Then consider which of the following types of digital products you will create in your project, and complete each section of the form that is applicable.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

Complete this section if your project will create digital content, resources, or assets. These include both digitized and born-digital products created by individuals, project teams, or through community gatherings during your project. Examples include, but are not limited to, still images, audio files, moving images, microfilm, object inventories, object catalogs, artworks, books, posters, curricula, field books, maps, notebooks, scientific labels, metadata schema, charts, tables, drawings, workflows, and teacher toolkits. Your project may involve making these materials available through public or access-controlled websites, kiosks, or live or recorded programs.

SECTION III: SOFTWARE

Complete this section if your project will create software, including any source code, algorithms, applications, and digital tools plus the accompanying documentation created by you during your project.

SECTION IV: RESEARCH DATA

Complete this section if your project will create research data, including recorded factual information and supporting documentation, commonly accepted as relevant to validating research findings and to supporting scholarly publications.

SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS

A.1 We expect applicants seeking federal funds for developing or creating digital products to release these files under open-source licenses to maximize access and promote reuse. What will be the intellectual property status of the digital products (i.e., digital content, resources, or assets; software; research data) you intend to create? What ownership rights will your organization assert over the files you intend to create, and what conditions will you impose on their access and use? Who will hold the copyright(s)? Explain and justify your licensing selections. Identify and explain the license under which you will release the files (e.g., a non-restrictive license such as BSD, GNU, MIT, Creative Commons licenses; RightsStatements.org statements). 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.

The toolkit resulting from this project will be made available under Creative Commons Attribution-NonCommercial license (CC BY-NC) 4.0 (https://creativecommons.org/licenses/by-nc/4.0/), which will allow users to copy and share the toolkit and its components, and, importantly, to adapt components to their own contexts and share the outcomes.

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.

Ownership is governed by the University of Maryland Intellectual Property Policy (https://president.umd.edu/sites/president.umd.edu/files/files/documents/policies/IV-320A.pdf). According to the Creative Commons Attribution-NonCommercial license, there will be no conditions on access and use, except that the toolkit is attributed and there are no commercial derivatives. This condition will not limit access to the toolkit by the stakeholders or intended beneficiaries of this project or the public at large. The toolkit will be made freely and openly available in an open-acess repository.

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.

N/A

SECTION II: 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 the format(s) you will use.

The central resource produced by this project will be a toolkit, expected to comprise mainly of text and image documentation. The specific formats will be govered by the research and are not determined, but the toolkit is likely to be made available in two formats: as an interactive web site, and as an information package containing multiple documents for download, sharing, editing, and repository deposit.

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

Standard document-creation software such as MS Office and Google Docs
Workflow visualization software (TBD)

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG, OBJ, DOC, PDF) you plan to use. If digitizing content, describe the quality standards (e.g., resolution, sampling rate, pixel dimensions) you will use for the files you will create.

DOC, JPG, PDF

Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

The project is designed to include iterative rounds of evaluation of the toolkit and its components (primarily workflows) by case partners and the project team.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should 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).

The toolkit aims to be self-documenting and will include documentation about its creation and intended use. It will be deposited in the University of Maryland Institutional Repository for long-term, open-access storage (https://drum.lib.umd.edu/).

Metadata

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

N/A

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

N/A

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).

N/A

Access and Use

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, delivery enabled by IIIF specifications).

The toolkit will be made available as an openly accessible website and deposited into the University of Maryland's institutional repository for preservation and ongoing access. The toolkit and all project outcomes will be broadly disseminated through the project website and social media, journal publication, and professional and research conferences oriented toward library and archives practitioners and communities developing digital collections.

D.2. Provide the name(s) and URL(s) (Universal Resource Locator), DOI (Digital Object Identifier), or other persistent identifier for any examples of previous digital content, resources, or assets your organization has created.

https://drum.lib.umd.edu/handle/1903/1631

SECTION III: SOFTWARE

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.

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

Technical Information

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

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

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

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

B.5 Provide the name(s), URL(s), and/or code repository locations for examples of any previous software your organization has created.

Access and Use

C.1 Describe how you will make the software and source code available to the public and/or its intended users.

C.2 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL:

SECTION IV: RESEARCH DATA

As part of the federal government's commitment to increase access to federally funded research data, Section IV represents the Data Management Plan (DMP) for research proposals and should reflect data management, dissemination, and preservation best practices in the applicant's area of research appropriate to the data that the project will generate.

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

Through interviews, site visits, participant observation, and documentary analysis, this project will produce qualitative data including interview recordings and transcripts, participant-observation memos, site visit field notes, and documentary analysis. There are three central cases with approximately 6-20 participants per case. There are 5-10 mini-cases planned with 1-5 participants per case. There are 10-12 interviews planned with representatives from memory institutions associated with the central cases. Data will be collected in years 1 and 2 of the project as described in the schedule of completion.

It is necessary to keep the data confidential because participants are being asked to share information about their work practices, professional status, opinions, funding, teamwork, and future plans that could embarrass participants if taken out of context or compromise relationships with stakeholders in each project, and which contains ex

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?

A preliminary data collection for the three central cases has obtained IRB approval, included in supporting documents. The other methods in this study, including mini-cases and interviews with memory institutions, will require approval from the UMD IRB. IRB approval will be sought 2 months prior to the planned start of data collection (i.e., in the final months of year 1).

A.3 Will you collect any sensitive information? This may include personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information. If so, detail the specific steps you will take to protect the information while you prepare it for public release (e.g., anonymizing individual identifiers, data aggregation). If the data will not be released publicly, explain why the data cannot be shared due to the protection of privacy, confidentiality, security, intellectual property, and other rights or requirements.

The data will not be released publicly per the terms of our agreement with case partners and our IRB approval. Data produced in this project is confidential and, under terms of its collection, will be destroyed after research is complete. It is necessary to keep the data confidential to foster trust with participants. Quotations will be used anonymously.

A.4 What technical (hardware and/or software) requirements or dependencies would be necessary for understanding retrieving, displaying, processing, or otherwise reusing the data?

N/A

A.5 What documentation (e.g., consent agreements, data documentation, codebooks, metadata, and analytical and procedural information) will you capture or create along with the data? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the data it describes to enable future reuse?

The data will not be shared. Some documentation, including information about research and analytical methods and the codebook, will be shared via the project website and dissemination plan described.

A.6 What is your plan for managing, disseminating, and preserving data after the completion of the award-funded project?

The raw data will be destroyed. The coded data will be stored securely in the UMD Box environment until research is complete and disseminated. It will be destroyed after seven years.

A.7 Identify where you will deposit the data:

Name of repository:

N/A

URL:

N/A

A.8 When and how frequently will you review this data management plan? How will the implementation be monitored?

N/A