

Abstract

The Pennsylvania Academic Library Consortium, Inc. (PALCI), in full partnership with the Private Academic Library Network of Indiana (PALNI), representing 94 academic libraries in Indiana, Pennsylvania, New Jersey, West Virginia, and New York, requests an IMLS National Leadership Grant for Libraries for \$130,900 to fund a 12-month Infrastructures and Initiatives Project to explore, develop, and pilot the Hyku open source, multi-tenant, consortial institutional repository (IR) to deliver ultra-low cost hosting, discovery, and access to digital material.

This project will further build upon the open source Hyku IR developed by the well-established Samvera community, which was previously supported by two IMLS projects grants (“Hydra in a Box,” LG-70-15-0006-15, and “Bridge2Hyku,” LG-70-17-0217-17) to create the functionality and configuration options needed by consortia using this multi-tenant solution. Partnering consortia will scale the implementation of the Hyku IR across two consortia, and build their member libraries’ capacity to serve their communities with access to digital collections by reducing hosting and administration costs using shared infrastructure, while increasing likelihood of widespread adoption, allowing individual libraries to customize and brand the IR as their own.

This project will employ a newly formed Cross-Consortial Hyku Task Force, consisting of cross-functional staff representing diverse member libraries from each consortium to identify and define key functional requirements, feature prioritization, and workflows. PALNI and PALCI will contract with experienced Hyku application developers to further develop, enhance and configure the software for scalable use across multiple consortia. Lastly this project will focus a significant effort on building required governance structures, developing cost recovery/business models, and creating sustainable administrative workflows and best practices. Project code will be shared back with the community under a permissive open source software license, and results will be reported regularly to the institutional repository and consortial communities.

This project will transform the open source institutional repository space by offering a needed multi-tenant alternative to today’s limited commercial and open source offerings, and will push traditional consortial boundaries to provide a highly scalable, easily adoptable, and sustainable IR service that ultimately supports IMLS goals to build capacity, increase public access to information, and support lifelong learning.

Project Narrative

Introduction

The Pennsylvania Academic Library Consortium, Inc. (PALCI), in full partnership with the Private Academic Library Network of Indiana (PALNI), representing 94 academic libraries in Indiana, Pennsylvania, New Jersey, West Virginia, and New York, requests an IMLS National Leadership Grant for Libraries for \$130,900 to fund a 12-month Infrastructures and Initiatives Project to explore, develop, and pilot the Hyku open source, multi-tenant, consortial institutional repository (IR) to deliver ultra-low cost hosting, discovery, and access to digital material.

This project will build upon the open source Hyku IR developed by the well-established Samvera community, previously supported by two IMLS projects grants (“Hydra in a Box,” LG-70-15-0006-15, and “Bridge2Hyku,” LG-70-17-0217-17), to create the functionality and configuration options needed by consortia to use this multi-tenant solution. Partnering consortia will scale the implementation of the Hyku IR to build their member libraries’ capacity to serve their communities with access to digital collections by reducing hosting and administration costs using shared infrastructure and increasing the likelihood of widespread adoption by allowing individual libraries to customize and brand the IR as their own. This project will push traditional consortial boundaries to provide a highly scalable, easily adoptable, and sustainable IR service that ultimately supports IMLS goals to build capacity, increase public access to information, and support lifelong learning.

Statement of National Need

Institutional Repositories are Critical to Scholarly Communications & Access

Institutional repositories (IRs) and digital asset management systems are critical to the life cycle of scholarly communications and the digital preservation of our cultural heritage, providing key preservation and access points to important research, special collections, and other types of information produced in our organizations. Academic libraries often use institutional repositories as a means for controlling and disseminating the university’s digital assets and demonstrating value to research communities and the greater public through these outputs. Repositories have been around since the 1990s and in 2003, Clifford Lynch, Director of the Coalition for Networked Information (CNI) described repositories as essential information architecture for universities. Lynch stated, “The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content, and also scaling up beyond ad-hoc alliances,

partnerships, and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium.”¹

Libraries have largely embraced their role in responsibility for this process, and many have a long history of providing IR services to offer storage, preservation and dissemination solutions that increase patron access to information for traditional types of materials, including Electronic Theses and Dissertations (ETDs), pre-prints, and published research. Institutional repositories connect researchers, including faculty, students and staff, allowing them to share their work on a worldwide scale, and often function as alternative pathways to open access content and other locally produced works that may otherwise be difficult to find and access. Cultural heritage institutions rely heavily on such software to host and describe important local collections, and provide access to digital objects that would otherwise be largely hidden from the world.

Today, libraries and consortia are increasingly looking to institutional repositories to offer storage and access points for new types of content such as research data sets and Open Educational Resources. These new types of IR content require the development of standardized metadata, description, and discovery. A number of communities are working to develop collaborative, open, and standardized approaches to improve efficiency. Organizations such as DataCite and CrossRef allow libraries and institutions to register their assets and research materials, including data sets, with Digital Object Identifiers (DOIs). These identifiers provide common metadata descriptions through persistent URLs, aiding researchers in locating and citing information.

A movement toward supporting Open Educational Resources (OER) is also growing among library consortia in the United States seeking to reduce the cost of textbooks and increase affordability of education for students. To that end, and, libraries are seeking ways to store and connect researchers and faculty to these locally-produced materials. Organizations such as the Open Textbook Network serve as referatories, pointing to curated open educational resources, which are often stored in a library’s institutional repository architecture. However, libraries need affordable IR solutions that work well with the open nature and iterative workflows associated with this type of content in order to adequately manage this emerging content type and make them available via OER aggregation services.

Proprietary and Open Source Options Fail to Meet Library Needs

Since the late 1990s libraries have been doing institutional repository work primarily on an individual library-by-library scale. A variety of commercial and open source solutions were developed and remain popular today, including BePress, CONTENTdm, DSpace, TIND, and Islandora, among others, which provide single tenant architecture, allowing individual libraries

¹ Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. *Libraries and the Academy* 3(2), 327-336. Johns Hopkins University Press.

to procure hosted IR services. And, for those with the technical capacity and staff expertise, options exist to host an open source locally administered and customized experience for their library community in a single repository.

Commercial repository options are often appealing as out-of-the-box solutions, and some include hosting at relatively low costs. However, low-cost options such as CONTENTdm are limited in the scope of services and ability to meet changing user needs. CONTENTdm has focused on the digital display of visual cultural heritage assets, and while excelling at storing and displaying a single content type, it does not provide institutional repository services as an integrated solution across multiple content formats and types.

Full-fledged commercial options on the other hand, offer a high degree of integration attractive to research-heavy institutions. Commercial offerings, such as Elsevier's BePress, have adapted workflows to well-support modern research activities, but these solutions are costly, greatly limiting the ability for many libraries, especially those with less resourced and underserved user populations, to take advantage of such services. In addition, this proprietary software is locked down and cannot be adjusted by libraries to meet evolving needs or collaborative workflows. These highly integrated solutions also raise other concerns related to commercial publisher control of the research community's data and content. A well-known movement documented among academic libraries known as "BePrexit" began in 2017 as libraries voiced their concern about the consolidation of BePress institutional repository services by Elsevier, and libraries, like the University of Pennsylvania responded by issuing statements and seeking alternative tools to meet institutional needs.²

There has been a shift toward adoption of open source IR solutions (DSpace, Hydra, Fedora, Islandora) in recent years, which offer flexibility, potential for innovation based on user needs, and vendor/publisher-neutral independence, as well as some degree of cost containment from fully integrated commercial options. In 2017, David Lewis published his popular 2.5% Commitment, urging libraries to dedicate 2.5% of their budgets to fund open access efforts, including the development of institutional repositories. These open source solutions provide varying levels of functionality to individual libraries, and many offer robust user communities that support these tools. But while there are a variety of single tenant open source options available, even large research universities fully capable of hosting their own individual solutions recognize the administrative overhead and difficulties in creating and maintaining sustainable local IR systems and workflows.

In addition, libraries are increasingly concerned about the siloing effect of individual single tenant institutional repository solutions lacking unified search and retrieval options. It is possible for an individual single tenant IR to be shared across multiple institutions, but not without losing

²<https://beprexit.wordpress.com/official-statement/>

important functionality and the capacity to effectively customize branding, workflows, and metadata templates to meet local needs. Single tenant solutions also require duplicative routine activities, such as software updates, staff support, hosting maintenance, and administrative set-up. Current single tenant open source systems have lacked the architecture and configuration options necessary for efficient and cost-effective services at consortial scale.

New, Affordable and Scalable Solutions are Required

Significant budget challenges, limitations of existing software, and consolidation of commercial services and software leave libraries seeking new solutions. Increasingly these libraries are looking to consortia to scale IR services and expand their ability to provide storage of and access to digital collections. In a survey of the Private Academic Library Network of Indiana (PALNI) supported libraries, over 70% did not have an IR system in place, and more than 65% were interested in participating in a collaborative IR system. Budget constraints, staffing limitations, and insufficient technological support were the greatest impediments identified in the survey. A PALNI environmental scan identified multiple IR software options; however, current open source single tenant systems lacked the large-scale configuration functionality necessary to share costs and infrastructure in a consortium setting.

Within the PALCI consortium, similar requests for IR services were made by member libraries seeking affordable IR solutions. Twenty percent of PALCI libraries (academic institutions with diverse student populations and institutional missions ranging from liberal arts-focus to research missions) indicated a desire to participate in a pilot and beta testing of the HykuDirect software, the IMLS-supported Hydra-in-a-Box project (LG-70-15-0006-15), with the goal of creating a turnkey digital repository solution³. This solution demonstrated great promise as an easy-to-use, hosted solution, but lacked key consortial configuration options and features necessary to effectively deliver service at scale that could be widely adopted by libraries.

Many library consortia are eager to explore meaningful partnerships in the IR space for greater shared benefit through cross-consortial collaboration. In 2017, PALCI and PALNI engaged members of the International Coalition of Library Consortia (ICOLC) in meetings at the American Library Association Midwinter and ICOLC Detroit conferences to discuss consortial IR needs where over thirty consortia expressed interest in collaborative IR efforts. PALNI developed a vision document describing the desired solution, and expressed interest in partnering with other groups in building out affordable infrastructure for these services. Consortia in attendance at these meetings with experience running IRs cautioned against the high degree of cost and technological support required to provide this service through existing software, yet none had explored shared infrastructure across consortial boundaries.

Hyku Development Builds Capacity, Aligns with IMLS Priorities

³ <https://duraspace.org/hykidirect-pilot-program-and-hyku-beta-testing-underway/>

The project proposed here builds on the success of previous IMLS-funded projects to develop the capacity of libraries and consortia to offer affordable, scalable IR services via the Hyku multi-tenant IR platform. Hyku represents a needed alternative for collaborative library infrastructure that will increase public access to information, and improve libraries' abilities to effectively serve as stewards of collections and materials that demonstrate the value of our institutions.

The PALCI and PALNI consortia have each made significant investments in time and resources toward developing low-cost IR services by directly funding and engaging in Hyku's development and contributing that development back to the larger community. In addition, each consortium is an active participant in the Samvera Hyku Interest Group, an open source community connecting the work in progress of other participating groups around the world, including work by the British Library⁴. This project leverages the investments already made by other organizations and IMLS, and the large library networks of PALCI and PALNI to offer significant experience and expertise in deep collaboration, and a strong desire to build capacity for many consortia across the United States and beyond.

Project Design

In Fall 2017, PALCI piloted a scaled-back, hosted version of Hyku, the only available open source, multi-tenant IR software. The HykuDirect pilot was part of a grant project run by Stanford, DuraSpace, and Digital Public Library of America (DPLA) and funded by IMLS (LG-70-15-0006-15). Tests showed the software was promising, but additional development work was required to build on the success of this project. In early 2018, PALNI developed a Consortial IR Vision (<https://goo.gl/58HCuz>), which provides a graphical sketch of the desired multi-tenant environment, maps out specifications for several functional levels, and defines user management needs. This document has been shared and used widely across the consortial and library communities and has been effective in conveying the consortial IR vision to partners and service providers.

Since Spring 2017, PALCI and PALNI have been jointly exploring IR software options, and after having completed an extensive environmental scan and discussions with other consortia, identified Hyku to have the highest long-term potential as an affordable, scalable consortial solution. Notch 8, a web development firm with substantial Hyku expertise and a leader of and major contributor to the Samvera community, has been identified as the project's lead application developer to create a multi-tenant Hyku instance for shared use by our two consortia. PALNI and PALCI have begun actively developing a collaborative consortial prototype Hyku IR which will require further development, testing, and evaluation to move toward a production-ready service. The project has been initially scoped to focus on adding two new content types to

⁴ <https://www.bl.uk/press-releases/2018/july/shared-research-repository-announcement#>

the Hyku IR service: 1) Open Educational Resources (OERs), and 2) Electronic Theses and Dissertations (ETDs), with expanded options available in future project phases.

Performance Measures & Outcomes - The success of this project will be measured based on our ability to: 1. contribute a fully-featured, multi-tenant, open source IR tool to consortial communities, enhancing library consortial ability to provide open access materials while containing costs; 2. develop a model for multi-institution/consortium collaboration and shared infrastructure that is easily adoptable; and 3. further grow the Hyku community by engaging library subject matter experts in its governance.

Target Audience - The project's initial audience will be member libraries from the PALCI and PALNI communities, with a much wider, secondary audience of all library consortia that are considering institutional repository services for their memberships. The growing group of libraries and consortia engaged in supporting workflows, interoperability of, and access to Open Educational Resources are also a key target audience as this project aims to further support their work with cost effective tools.

Project Schedule Overview & Work Plan

Phase 1: Cross-Consortial Hyku Task Force: Functional Requirements, Metadata Template Definitions, Workflows, & Feature Prioritization (July 2019 - June 2020)

The formation of this strategic group will occur early in the project. The group will hold regular meetings throughout to represent PALCI and PALNI member libraries with local staff experts who will be charged with specifying key functionality, identifying features and enhancements, outlining shared/unique needs across consortia, and leading the pilot and assessment. The Task Force will meet at least one time in-person, in addition to regular web-based meetings to achieve their charge. Key milestones in Phase 1 include:

- Needs assessment for use cases, workflows, and functionality
- Organize working groups for ETD and OER worktype and workflow development
- Collaboration with external advisors and related organizations for worktype feedback, such as ISKME, SPARC, Open Textbook Network, the British Library, and others
- Exploration of consortium scale DOI services (e.g., DataCite, CrossRef) to aid in discoverability and access to research and content stored in the IR
- In-person meeting

Phase 2: Develop and Enhance Hyku's Multi-Tenant Software Development for Collaborative Consortial Use (July 2019 - June 2020)

The Hyku software development will focus on expanding available content types, discovery, configuration, and metadata schemas in support of shared consortial IR services, as well as the ability to efficiently manage individual institutions' branding templates. Software development

and documentation will be provided back to the community under a permissive Apache 2.0 open source license. Key milestones in Phase 2 include:

- Theming and branding development
- Refining user permissions and roles for multi-tenant administration
- Building collaborative workflows
- Multi-tenant viewable works
- Cross collection searching; enhance data exports for improved discovery
- ETD worktype implementation
- OER worktype implementation
- Define worktype versioning workflows
- Integration with external Hyku development efforts

Phase 3: Hyku Software Pilot and Communication of Project Results (October 2019 - June 2020)
Select PALNI and PALCI libraries will pilot the software to test for scalability, assess sustainability and ease of adoption, and communicate results by making reports to the consortial and institutional repository community. Notch 8 will host the repository and provide backup of the associated files and metadata records during this phase. A future project phase will include the creation and implementation of a robust digital preservation plan for a production ready consortial IR.

Phase 3 will focus heavily on communication of the project's progress and results. Conferences targeted for reporting to the community include: Samvera Connect, Code4Lib, Open Repositories, ICOLC, Charleston, and ALA, as well as meetings related to Open Educational Resource networks (e.g., OpenEd). These and other reports will describe the product and how to use it, but also lessons learned on building the collaborative structure and governance, adoption narratives, sustainability, and costs/business models. The project code will be distributed to the community via GitHub. Key milestones in Phase 3 include:

- Early development testing
- Pilot phase
- Project reporting, documentation, and training
- Build out sustainability/governance/business models
- Outreach & communications
- Contribute code and development efforts back to Hyku/Samvera community

Risks - Risks associated with this project are minimal and similar to other software projects implementing existing open source software solutions. Risks have been mitigated through creation of a MOU between PALNI and PALCI, and through previous testing of the software (HykuDirect pilot and initial self-funded Hyku development phases, demonstrating the project's feasibility). Other identified risks may include:

- Major marketplace changes, e.g., new availability of a widely affordable commercial solution could take market share away from those interested in the Hyku open source solution. Mitigation strategy: Actively communicate and encourage engagement with the consortium community to explore current and newly available options. Build transparent, sustainable business models for supporting shared services to increase partnerships within the consortium community.
- Major changes to project personnel, e.g., loss of key project staff. Mitigation strategy: develop essential project documentation and expertise/responsibility among project planners.

Budget & Project Resources - The success of this project is dependent on the expertise of the Cross-Consortial Hyku Task Force, the project's contracted application developers, and the commitments made by the PALCI and PALNI consortia in developing shared infrastructure and services. Initial self-funded feasibility studies and project phases have established a basic collaborative consortial Hyku software implementation in preparation for further development and enhancement to create a minimally viable service.

Though not required, significant in-kind contributions are anticipated in the form of Cross-Consortial Hyku Task Force time, and initial project contributions for early project phases by PALCI and PALNI. The budget below is presented with a high degree of confidence based on previous experience, developer estimates and project specifications.

Web hosting, storage, and infrastructure for 12 months - \$14,000;

Contract for Hyku software development - \$50,000;

Project management and staffing salaries - \$25,000;

Travel supporting project governance, outreach & communication - \$25,000;

Cross-Consortial Hyku Task Force and stakeholder community meeting expenses - \$5,000;

Indirect costs at 10% rate - \$11,900;

Total Requested Grant Funds = \$130,900

Diversity Plan

The Collaborative Consortial IR project will engage a broad range of stakeholders across multiple consortia to produce a software that will serve the diverse communities of many libraries, representing academic library users of all types. Consortial IR software has the potential to greatly reduce the burden of increasing public access to research and other materials produced by our organizations, in addition to broadening the reach of special collections to populations who would otherwise not have access. This project is also scoped to focus on Open Educational Resources, aiding traditionally underserved populations by making available open course materials and supporting affordability in higher education. Lastly, care will be taken to

solicit and incorporate feedback from traditionally under-represented groups in each consortium with a particular focus on meeting or exceeding web accessibility standards. PALNI's newly formed consortial accessibility task force will be consulted as a source of expertise for this project.

National Impact

The success of this project will further cement libraries' place as central to delivery of key repository services to their communities, ensuring sustainable, affordable solutions for preservation and provision of access to digital content. This project will also enable our consortia and others to build on the development of the Hyku IR software for the good of the entire community. Potential beneficiaries of this project encompass thousands of libraries from academic library consortia across the world, in addition to public libraries and other organizations seeking to make digital content widely available. All code produced in this project will be made available via permissive open source licenses, allowing for maximum freedom to adopt and innovate on Hyku.

Additionally, widespread adoption of a multi-tenant IR solution will transform libraries' ability to share infrastructure nationally and beyond, and reduce reliance on commercial solutions. The Hyku community is engaged and growing, and our selected developer, Notch 8, is a major contributor to this group, working in coordination with other libraries and developers, including Ubiquity Press and the British Library. As part of this project, PALNI and PALCI will develop sustainable cost recovery and funding models, allowing for ongoing success and widespread adoption. These partnerships are key to success and will result in a long-term sustainable IR solution for libraries.

The national impact of this project aligns with IMLS priorities by:

- 1) Building needed capacity and alternative solutions to improve public access to information
- 2) Exploring, enhancing, and scaling an easily adoptable multi-tenant IR solution previously developed with IMLS funds that will serve the greater consortium community
- 3) Developing partnerships across library communities to effectively sustain open source tools and services and contain costs
- 4) Enhancing the sustainability, interoperability, and accessibility of digital content and collections to provide long-term impact to and value for diverse and evolving user communities
- 5) Advancing realistic and sustainable approaches for the stewardship of new and complex content types or digital formats.
- 6) Enabling the ethical stewardship of diverse or culturally sensitive digital content and collections through inclusive collaborations, tools, and best practices

**Pennsylvania Academic Library Consortium, Inc. (PALCI), Scaling Up a Collaborative Consortial Institutional Repository,
March 2019 Proposal to IMLS**

Scaling up a Collaborative Consortial Institutional Repository (CC-IR)

Project Schedule

ACTIVITY	Jul-19	Aug-19	Sep-19	Oct-19	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20
Project setup: Hire project coordinator, review project plan	█	█	█									
Phase 1: Cross-Consortial Hyku Task Force: Functional Requirements, Metadata Template Definitions, Workflows, & Feature Prioritization												
-Form and convene cross-consortial Hyku pilot task force	█											
-Needs assessment for use cases, workflows, and functionality	█	█	█									
-Working groups for ETD and OER worktype development				█								
-Collaborate with external advisors for worktype feedback	█	█	█	█	█	█	█	█	█	█	█	█
-Exploration of consortium scale DOI services												
-In-person meeting				█								
Phase 2: Develop and Enhance Hyku's Multi-Tenant Software Development												
-Theming and branding development	█	█	█									
-Refining user permissions and roles for multi-tenant administration				█	█	█						
-Building collaborative workflows				█								
-Multi-tenant viewable works	█	█	█									
-Cross collection searching; enhance data exports for improved discovery				█	█	█	█	█	█	█	█	█
-ETD worktype implementation					█	█						
-OER worktype implementation								█	█			
-Define worktype versioning workflows				█	█	█	█	█	█	█	█	█
-Integration with external Hyku development efforts	█	█	█	█	█	█	█	█	█	█	█	█
Phase 3: Hyku Software Pilot and Communication of Project Results												
-Early development testing				█	█	█						
-Pilot phase							█	█	█	█	█	█
-Project reporting, documentation, and training										█	█	█
-Build out sustainability/governance/business models										█	█	█
-Outreach & communications										█	█	█
-Contribute code and development efforts back to Hyku/Samvera community										█	█	█



DIGITAL PRODUCT FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (e.g., 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. 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

All applications must include a Digital Product Form.

- 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.

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.

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.

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 the format(s) you will use.

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.

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).

B. Workflow and Asset Maintenance/Preservation

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

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).

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).

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

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).

D. 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).

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.

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.

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.

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.

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) and URL(s) for examples of any previous software your organization has created.

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.

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

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:

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.

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.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).

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.

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).

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?

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

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

Name of repository:

URL:

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