Many cultural heritage institutions have used a variety of solutions to preserve important digital content either generated or collected by their constituencies. Software repository systems may provide some preservation functionality, but they pose risk factors such as lack of geographic diversity, lack of technological diversity, and loss of data related to human activities and systems failures. Distributed digital preservation (DDP) systems, networks in geographically dispersed locations designed to perform preservation actions, aid in mitigating these risk factors. As institutions increase their local preservation capabilities, they often face challenges integrating local repository systems with DDP systems.

Northwestern University and Educopia will bring together a diverse group of digital preservation experts, practitioners, and service providers to collaboratively develop tools for local repository systems and DDP services to work together more effectively. This National Leadership Project Grant will build upon research conducted in 2017 by Northwestern University and the University of California San Diego on the challenges that cultural heritage organizations encountered when trying to integrate local repositories with DDP systems, as part of an IMLS planning grant (LG-72-16-0135-16). The planning grant research resulted in the following recommendations: the creation of a decision-making curation toolkit for choosing materials to send to DDP systems; the development of a common BagIt profile that cultural heritage organizations could use to submit materials to any DDP system; and the development of a tool that would enable organizations to track materials that they have sent to DDP systems.

In the first year of this two-year National Leadership Project Grant, teams dedicated to creating the curation toolkit and to developing the shared BagIt profile will meet several times in Evanston, IL. Once the BagIt profile is developed and agreed upon, software developers at two local repository systems, Fedora and DSpace, and at two DDP services, LOCKSS and the Academic Preservation Trust (APTrust), will work to implement the new profile in their systems. After the development work occurs, users of these repository and DDP services will test the new tools and workflows, and any changes will be made by the software developers based on user feedback at the end of the two-year grant term. Finally, code written at APTrust as part of this grant will be shared with the Digital Preservation Network and distributed through the nodes of that network to promote interoperability among more DDP services after the grant is complete.

This project is targeted at those engaged in digital preservation activities at various levels and at any type of organization, as well as at DDP service providers. By promoting greater efficiency and interoperability in digital preservation workflows and systems, the work conducted under this project grant will build on the IMLS National Digital Platform. The outcomes of this project grant will reduce barriers to effective digital preservation: the curation toolkit will promote efficiency in selecting materials for distributed digital preservation, while the shared BagIt profile and software development work required to implement it will bridge gaps between local repository and DDP infrastructure. All of the work that will be conducted under this project grant will be highly collaborative and in the spirit of the Digital Preservation Declaration of Shared Values, which promotes unity based on "commitment to preserve the cultural, intellectual, scientific and academic record for current and future generations."

This National Leadership Project Grant will build upon research conducted in 2017 by Northwestern University and the University of California San Diego on the challenges that cultural heritage organizations encountered when trying to integrate local repositories with distributed digital preservation (DDP) systems, as part of an IMLS planning grant (LG-72-16-0135-16). The goals of this grant are: 1) to create a curation toolkit that will assist cultural heritage organizations in choosing materials to send to DDP systems; 2) to create a shared BagIt profile to increase interoperability between repositories and DDP services; 3) to develop DDP and local repository systems to accept the new BagIt profile. The funds requested from the IMLS to complete this work in two years are: $249,607. The ultimate outcome of this grant is to improve the management and preservation of content and collections nationwide at cultural heritage organizations of various types and sizes through better tracking of content from local repositories to distributed digital preservation systems.

Statement of National Need
Many cultural heritage institutions have used a variety of solutions to preserve important digital content either generated or collected by their constituencies. Local software repository systems, such as Fedora and DSpace, may provide some preservation functionality, but they pose risk factors such as lack of geographic diversity, lack of technological diversity, and loss of data related to human activities and systems failures. Distributed digital preservation (DDP) systems, networks in geographically-dispersed locations designed to perform preservation actions, aid in mitigating these risk factors. As institutions increase their local preservation capabilities, they often face challenges integrating local repository systems with DDP systems.

In 2017, as part of an IMLS planning grant (LG-72-16-0135-16), Northwestern University Libraries and the University of California San Diego Library conducted research into the challenges that cultural heritage organizations encountered when trying to integrate local repositories with DDP systems. By surveying cultural heritage organizations of different types and sizes about their digital preservation practices, the researchers gathered information about gaps in existing technologies and infrastructure that are preventing efficient and sustainable digital preservation in many organizations. Upon analyzing the 170 survey responses and twelve follow-up interviews, the researchers proposed three technical recommendations that would address challenges in integrating local repositories with DDP systems: the creation of a decision-making curation toolkit for choosing materials to send to DDP systems; the development of a common BagIt profile that cultural heritage organizations could use to submit materials to any DDP system; and the development of a dashboard-type tool that would enable organizations to track materials that they have sent to DDP systems.[1] This two-year National Leadership Project Grant will build upon this IMLS-funded research by bringing together a diverse group of digital preservation experts, practitioners, and service providers to collaboratively develop tools for local repository systems and DDP services to work together more effectively. This work will be useful for organizations of all shapes and sizes, and will be guided by the goals of the IMLS National Digital Platform to "[bridge] the gaps between disparate pieces of the existing digital infrastructure" to promote increased efficiencies, cost savings, access, and services.[2]

---

[1] Evviva Weinraub et al., "Beyond the Repository: Integrating Local Preservation Systems with National Distribution Services," 2018, doi:10.21985/N28M2Z, 34-35.

[2] Institute of Museum and Library Services, "National Digital Platform," https://www.imls.gov/issues/national-issues/national-digital-platform

The first outcome of this grant, the curation toolkit, will be targeted at cultural heritage organizations who have committed to long-term preservation of their digital assets and wish to mitigate risk by distributing those assets into DDP systems. A significant challenge facing this group is the selection of materials to send to DDP systems. Because the cost of distributed digital preservation is dependent on the amount of content preserved, many organizations do not have the resources to store all of their digital content in DDP systems. Therefore, they need to select a subset of their content, but research in the planning grant indicates that few have developed policies or procedures for doing so.[3] Research also shows that there is little consensus on the criteria that should guide selection for sending materials to DDP systems: though survey respondents ranked "mandate" and "intrinsic value" as important selection criteria, responses on this topic reveal that respondents do not agree on the relative importance of the various criteria. The curation toolkit created in this project grant will identify various criteria that should be considered in selecting materials for DDP services and provide guidance on the decision-making process. The toolkit will benefit cultural heritage organizations engaged in digital preservation by making the selection process more efficient and ensuring that valuable or significant materials are preserved in DDP systems.

This next outcome of this grant, the shared BagIt profile, will also address needs expressed in the planning grant research. Cultural heritage organizations who participated in the research indicated that a lack of interoperability between digital preservation tools and systems was one of their biggest challenges, and generally expressed frustration with the lack of integration between systems.[4] A packaging format, such as a BagIt profile that can be understood by many systems, including local repositories and DDP services, would increase interoperability between pieces of the digital preservation infrastructure. Currently, different systems use different BagIt profiles or different packaging formats. This project grant will bring together representatives from cultural heritage organizations as well as local repository and DDP services to create a shared profile and to ensure that it will be used going forward across the digital preservation landscape of repositories and systems. This will reduce the number of workarounds required to integrate different parts of a digital preservation workflow.

The shared BagIt profile developed in this project grant will also benefit producers and developers of digital preservation services, such as local repository and DDP systems. Representatives from a variety of these types of organizations participated in the research conducted in the planning grant, and will collaborate further as part of this project grant. Software developers from the Fedora and DSpace projects, open source local repository systems, will implement the shared BagIt profile so that organizations using Fedora or DSpace-based repositories can export bags in the shared format. Similarly, developers from the LOCKSS Program and the Academic Preservation Trust (APTrust) DDP services will enhance their systems to accept bags in the shared format. LOCKSS developers will extend their metadata database schema to support metadata from the new BagIt profile, and APTrust developers will expose information from the bags so that cultural heritage organizations can more easily track their content. Code written at APTrust as part of this grant with the Digital Preservation Network (DPN), which will make it available to all the nodes in the network to promote interoperability among more DDP services after the grant is complete. Although one of the recommendations

---

[3] Weinraub et al., "Beyond the Repository," 26.

[4] Ibid., 30.

from the planning grant was to create a dashboard-type feature for tracking content that would be shared among DDP systems, this was deemed to be outside the scope of this project grant. Despite this, the development work at multiple local repository and DDP systems for this grant will result in increased linkages between systems, and will provide a foundation for other producers and developers of digital preservation services to build upon. For example, AVP will update their Exactly data transfer application to support the new BagIt profile developed in this grant.

By promoting greater efficiency and interoperability in digital preservation workflows and systems, the work conducted under this project grant will build on the IMLS National Digital Platform. Research conducted for the planning grant revealed significant barriers to effective digital preservation in a complex ecosystem consisting of a variety of tools, software, and services. The outcomes of this project grant will reduce these barriers, which spread across cultural heritage organizations of all shapes and sizes: the curation toolkit will promote efficiency in selecting materials for distributed digital preservation, while the shared BagIt profile and software development work required to implement it will bridge gaps between local repository and DDP infrastructure. All of the work that will be conducted under this project grant will be highly collaborative and in the spirit of the Digital Preservation Declaration of Shared Values, which promotes unity based on "commitment to preserve the cultural, intellectual, scientific and academic record for current and future generations."

Project Design

The goals of this grant are:
- to create a curation toolkit that will assist cultural heritage organizations in choosing materials to send to distributed digital preservation (DDP) systems
- to create a shared BagIt profile to increase interoperability between repositories and DDP services
- for the Fedora and DSpace repository systems to implement the shared profile in their BagIt export feature
- for LOCKSS and APTrust to update their systems to be able to accept the shared BagIt profile; LOCKSS will extend their metadata database schema to support metadata from the profile and APTrust will enhance their API to expose tags from the profile, thereby allowing users to generate reports on content they submitted to the DDP provider

The potential risk of a large multi-institutional collaborative project is that there will be differing opinions, but participants have already committed to build consensus and agreement in the digital preservation field as evidenced by the Digital Preservation Declaration of Shared Values. Additionally, with a large multi-institutional collaborative project there is a risk that participants may not fulfill their obligations to get work done, but the grant team strategically placed the in-person meetings to mitigate the risk of participants losing focus offline. Another risk is that there may be unknowns in the software development work that could impact the timeline. However, the proposed work plan includes time for user testing and additional iteration on development which will mitigate this risk. Lastly, a risk for a future outcome could be that any API specification that is crafted from the work of this grant could end up not matching real-world requirements for implementation and may need refinement later on. Any API that does get

developed as a result of the shared BagIt profile work should be flexible enough to account for future developments in metadata standards.

These project activities for this grant project are directly informed by the outcome of the previous planning grant, which included a literature review about distributed digital preservation systems, a survey that generated 170 responses from multiple types of institutions, twelve follow-up interviews with survey respondents, and data analysis with participation from an Advisory Board that culminated in a final report with three recommendations for increasing interoperability of digital preservation systems.

The activities for this grant will consist of the following:
- In collaboration with the National Digital Stewardship Alliance (NDSA), a working group will create a decision-making toolkit for choosing materials to send to DDP systems. (Months 1 – 10)
- A second working group will establish a shared BagIt profile that will increase interoperability between DDP systems. (Months 1 – 10)
- Fedora and DSpace (part of the second working group) will implement the shared BagIt profile in the export feature in each system. (Months 11-15; 21-23 if needed based on testing results)
- DDP systems (LOCKSS and APTrust, part of the second working group) will enhance their systems to be able to accept the shared BagIt profile; LOCKSS will also extend their metadata database schema to support metadata from the profile and APTrust will enhance their API to expose tags from the profile to allow reporting. (Months 11-15; 21-23 if needed based on testing results)
- At least four cultural heritage organizations will test exporting bags with the new shared profile from Fedora and DSpace and send them to their DDP systems (Months 16-20)

Travel for activities above:
- Travel for members of the NDSA working group to meet in Evanston, IL during months 3 and 7
- Travel for representatives from identified repository and DDP systems and the user community to meet in Evanston, IL during months 2, 5, 10
- Travel for appropriate representative to present work at three relevant preservation oriented conferences

The design of this grant derives from research conducted under an IMLS planning grant (LG-72-16-0135-16) that surveyed cultural heritage organizations about their digital preservation practices and use of DDP services. The grant team for that project included DDP service providers and experts from the digital preservation field, who recommended the creation of a curation toolkit and shared BagIt profile based on needs expressed by the 170 organizations who responded to the survey. This grant is similarly designed to bring together digital preservation experts, including DDP service providers and users of these services, to develop tools based on the research conducted as part of the planning grant. Participants in this grant will meet several times in person and conduct some collaborative work virtually. The in-person meetings will be

essential for facilitating debate and discussion and encouraging focus to complete the deliverables.

This project is targeted at repository and DDP service providers and those engaged in digital preservation activities at various levels and at any type of cultural heritage organization. The grant work will involve a wide range of people in the digital preservation field collaborating on the creation of a curation toolkit, development of a shared BagIt profile, and implementation of the shared profile in repository and DDP software. The toolkit and profile will be openly accessible after the grant is complete so that they can be widely used in the digital preservation field. Laura Alagna, Digital Preservation Librarian at Northwestern University Libraries, will facilitate the curation toolkit meetings, and will function as the overall project manager to ensure that all deliverables are met, including the software development work. Sibyl Schaefer, the Chronopolis Program Manager at USCD, will facilitate the BagIt meetings. Milestones for each in-person meeting will be set in the first meeting for both teams so that progress can be tracked. Lastly, Educopia staff will be responsible for the administrative duties surrounding the travel to in-person meetings.

Two teams will be dedicated to this project over the course of the two-year grant. One team will be dedicated to creating the curation toolkit and the second will work on the shared BagIt profile. The curation toolkit team will meet twice in person in Evanston, IL at Northwestern University Libraries to create the toolkit, and they will also hold regular calls in order to meet their goals. The curation toolkit team will consider research on current curation practices, such as the data collected in the planning grant, and will develop a set of criteria that will be used to evaluate content for digital preservation needs. The aim of the toolkit will be to assist digital preservation practitioners in deciding which content should be selected for the level of digital preservation that is provided by DDP services. The end product will take the form of a workflow diagram and textual document so that it will be sustainable. Additionally, the ends product will be hosted through Open Science Framework's NDSA portal. This work will take approximately one year to collect and write up appropriately.

The second team will consist of stakeholders from the identified DDP projects, as well as representatives from the user community who have been doing work in this area, and will meet three times in Evanston, IL at Northwestern University Libraries to agree upon a shared BagIt profile. Funding from this grant will also be used to pay for local development work within the various repository and DDP systems to enhance their services to accept the new profile. For both the development work on the local repository and DDP side, the project manager will facilitate weekly check-ins to make sure the development is progressing. The project manager will also include the organizations who will test the workflow for timely feedback throughout the development cycle, similar to many agile software development methodologies. Part of the development and testing work will include determining the workflow end-to-end: the ingest to and export from a local repository, to the ingest on the side of the DDP system, and the exposing of the data back to the submitter. Therefore, the testing will consist of sending bags conforming to the new profile from a test repository instance to the institution's DDP service provider.

Lastly, funds from this grant will also be used towards domestic travel to attend three conferences and meetings in North America that are directly relevant to the dissemination of this work, such as PASIG, SAA, Open Repositories, DLF, DigiPres, or iPres.

We are requesting funding ($72,215) for travel for twenty members of the two teams to attend face-to-face meetings hosted at Northwestern University. These funds will cover airfare, hotel, ground transportation, and per diem for all members of the teams. Funding is also being requested ($113,480) for development work at the local repository systems, Fedora and DSpace, and at the DDP providers, LOCKSS and APTrust, to implement the new BagIt profile. Travel funds ($6,216) are requested to attend relevant conferences to promote the work and share the results widely with the digital preservation community. The travel and conference funding will be a subcontract with Educopia for an estimated total of $86,274 including their facilities and administration costs. The total funds requested for this project is $249,607 (direct cost: $199,754; indirect costs: $49,853).

Personnel
Evviva Weinraub (Project Director), Assistant University Librarian for Collections Services & Technologies, Northwestern University

BagIt Team
- Carolyn Caizzi, Head of Repository & Digital Curation, Northwestern University
- Laura Alagna - Digital Preservation Librarian, Northwestern University (Project Manager)
- Sibyl Schaefer - Chronopolis Program Manager/Digital Preservation Analyst, University of California San Diego - (Facilitator)
- Andrew Woods - Technical Lead for Fedora, DuraSpace
- Bill Branan - Services Technical Director, DuraSpace
- Tim Donohue - Technical Lead for DSpace and DSpaceDirect, DuraSpace
- Dave Pcolar - CTO, Digital Preservation Network (DPN)
- Michael Ritter - Developer, Chronopolis & DPN (University of Maryland)
- Rosalyn Metz - Director of Library Technology and Digital Strategies, Emory University
- Sam Meister - Preservation Communities Manager, Educopia Institute
- Katherine Skinner - Executive Director, Educopia Institute
- Justin Simpson - Director of Archivematica Technical Services
- Bertram Lyons - Senior Consultant, AVPreserve
- Thib Guicherd-Callin - LOCKSS Technical Manager
- John Kunze - Identifier Systems Architect, California Digital Library
- Andrew Diamond - Senior Software Engineer, APTrust
- Tim Dilauro - Data Archivist, Johns Hopkins University

Curation Toolkit Team
- Laura Alagna - Digital Preservation Librarian, Northwestern University (Facilitator)
- Lauren Work - Digital Preservation Librarian, University of Virginia and NDSA Content Working Group co-chair
- NDSA Working Group - 4 people; to be selected once grant is awarded by the chair. This group will consist of digital preservation practitioners who have

experience sending materials to DDP systems. Efforts will be made to ensure that this group represents a variety of cultural heritage organizations.

The grant teams will make their deliverables, the toolkit and the BagIt profile, publicly available by the end of month 10. The teams will also use the appropriate organizational channels, like NDSA working group calls and blog posts on DDP provider websites, to share and promote their work. The completion of the BagIt profile is necessary for the remainder of the grant work. The testing of the workflow from a local repository to a DDP system at several cultural heritage organizations during months 16-20 will indicate the success of both the profile and the repository and DDP development work that takes place between months 11-15.

There are several ways that the work conducted as part of this project grant will be sustained. One outcome is that the digital preservation community can build upon this grant work in multiple ways. Because the NDSA Content Interest Group will be involved in creating and hosting the curation toolkit, future members of this group can update or adjust the toolkit as needed. The shared BagIt profile and related development work are also positioned so that the digital preservation community can continue these efforts. Beyond the life of this project grant, BagIt profiles used for export from local repository systems will represent a point of collaboration as the recommended community export package. Additionally, this shared BagIt profile is the first step in committing to create a common API specification that could then be used to track content in multiple DDP systems. A defined common API specification could lead to the dashboard that cultural heritage organizations indicated would be useful during the planning grant research. Creating such an API specification may take multiple years and was deemed outside the scope of this grant work, but our assumption is that the partners involved in creating the shared BagIt profile will naturally coalesce and commit to generating an API specification in the near future. Software development work done at APTrust as part of this grant will be shared with DPN, and distributed through the nodes of that network to promote interoperability among more DDP services after the grant is complete. Another way that the digital preservation community could build upon this grant work is that the increased interoperability between local repositories and DDP systems would position cultural heritage organizations to more easily test their disaster recovery plans by retrieving their data from DDP systems. Finally, the deliverables of this grant will be made openly available, so other third party organizations would be able to build upon this work in other ways. All of this grant work is in the spirit of furthering the sentiment of the [Digital Preservation Declaration of Shared Values](#), which embodies a commitment to sustaining digital cultural heritage assets and the software, tools, and specifications that support digital preservation activities on those assets.

Diversity Plan
The research conducted as part of the planning grant reached a diverse group of cultural heritage organizations. 170 different organizations responded to the initial survey, representing 35 states and the District of Columbia, as well as ten countries outside of the United States. The majority of respondents self-identified their organizations as academic institutions, though a significant number of participants also identified themselves as representing independent archives, government organizations, museums, nonprofit organizations, and public libraries.
The planning grant also included in-depth follow up interviews with twelve survey respondents; the interviewees were selected with geographic and organizational diversity in mind, and

consisted of public and private universities of varying sizes, a museum, and a government organization. The planning grant research revealed the digital preservation needs of this wide variety of organizations, and the outcomes of this National Leadership Project Grant are designed to address these diverse needs.

National Impact

This grant project will lead to systemic change within the digital preservation community as repository software creators, DDP service providers, and cultural heritage organizations agree to a shared BagIt profile. This profile will enable local repositories to track their content in DDP systems; it would also allow institutions to transfer and keep track of bags of digital objects from one DDP provider to another as digital preservation infrastructure and needs change and evolve. The shared BagIt profile will provide greater flexibility for organizations that employ DDP services, as well as allowing organizations who do not engage with DDP services, but who use BagIt locally to manage content in a standardized way. This work will also have an impact on an international level, as organizations begin to work through how to legally store digital objects across national borders. By creating and implementing these standards, this grant work will pave the way for more seamlessly sharing and storing content regardless of its origin or storage locations.

The creation of a curation toolkit will help cultural heritage organizations make informed decisions about which objects in their local repositories require different levels of digital preservation activities. Research conducted as part of the planning grant demonstrated that many cultural heritage organizations engaged in digital preservation struggled with identifying and selecting content to send to DDP systems. The toolkit will guide these organizations by making the selection process more efficient and ensuring that the most valuable or significant materials across the nation are preserved in DDP systems. Additionally, as mentioned above, the shared BagIt profile will provide all organizations using bags as part of their digital preservation activities a standard set of protocols that will be used with all DDP systems, thereby allowing for better interoperability and transferability of digital preservation objects across digital preservation platforms.

The curation toolkit will be openly accessible through the Open Science Framework and will be written so that organizations of diverse size and shape can utilize it in their daily workflows and adapt it if necessary. Furthermore, the local repository and DDP systems that are participating in this grant were selected based on the planning grant research, which indicated that they are the most widely-used systems among the cultural heritage organizations who responded to the survey. Involving the most common repository and DDP systems in this grant work ensures that the deliverables will be widely adaptable throughout the digital preservation community.

Outcomes from this grant project will be sustained in several ways. In terms of the curation toolkit, we anticipate that members of the NDSA Content Interest Group will continue adjusting and updating the toolkit since this group will be instrumental in its creation. The shared BagIt profile and related development work are also positioned so that the digital preservation community can continue these efforts. Research conducted as part of the planning grant showed that cultural heritage organizations engaged in digital preservation need a common dashboard to track and audit content they have submitted to DDP services. Such a dashboard or similar tool

could be created based on an open API specification that would expose information from bags in the shared BagIt profile. Creating such an API specification may take multiple years and was deemed outside the scope of this grant work, but many of the partners involved in creating the shared BagIt profile have indicated that they are interested in contributing to an API specification in the near future or doing other work to sustain the outcomes of this grant. For example, APTrust will share code written as part of this grant with DPN, which will make it available to all the nodes of the DPN network to promote interoperability among more DDP services after the grant is complete. In addition, the outcomes of this grant will be made openly available, various DDP providers could implement the shared BagIt profile in their own systems, and other third party organizations would be able to build upon this work as well.

The ultimate outcome of this grant is to improve the management of content and collections nationwide through better tracking of content from local repositories to distributed digital preservation systems. The work of this grant will improve the preservation of content and collections nationwide because the curation toolkit will allow for better selection of digital content. Additionally, the work of the grant will improve preservation of content and collections because these assets will be tracked more easily across local and national distributed systems. Though this grant will enable better data collection about the preservation of content and collections, it will not produce that data.

Success will be measured by the creation of a toolkit for selection and curation and a shared BagIt profile within the first year of the grant. Another benchmark for success will be the completion of the development work by repository and DDP software creators. During the testing period, at least four cultural heritage organizations will test the end-to-end workflow from a local repository system to a DDP service based on a testing framework to be decided upon prior in month 16. The success of these tests will also determine the success of the grant project as it will demonstrate that sharing a BagIt profile helps to improve the management of collections in tracking content from local systems to distributed ones.

Activity	Dec 18	Jan 19	Feb 19	Mar 19	Apr 19	May 19	Jun 19	Jul 19	Aug 19	Sep 19	Oct 19	Nov 19
BagIt Working Group		▓	▓	▓	▓	▓	▓	▓	▓	▓		
BagIt In-Person Meetings		▓			▓					▓		
Curation Working Group		▓	▓	▓	▓	▓	▓	▓	▓			
Curation In-Person Meetings			▓				▓					
Repository Software Development											▓	Cont.
DDP Development											▓	Cont.
University Testing of Development												
Iteration Phase												

Activity	Dec 19	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20	Jul 20	Aug 20	Sep 20	Oct 20	Nov 20
BagIt Working Group												
BagIt In-Person Meetings												
Curation Working Group												
Curation In-Person Meetings												
Repository Software Development	Cont. from Nov 19											
DDP Development	Cont. from Nov 19											
University Testing of Development												
Iteration Phase												

DIGITAL PRODUCT FORM

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

Instructions

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

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

Part I: Intellectual Property Rights and Permissions

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

The curation toolkit will be made openly accessible on NDSA's Open Science Framework's portal and will have a Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), allowing others to build upon the toolkit. Both repository and DDP systems are open source software projects and their code is available on Github, using aopen license terms like Apache Public license and BSD-3 Clause.

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.

Northwestern will not assert ownership and there will be no conditions on access except for attribution on the toolkit.

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.

Not applicable.

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

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

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.
The project will create a toolkit in the form of textual material, a BagItProfile, and edit existing software code.

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.
NDSA working group will create textual toolkit and release on Open Science Framework portal. BagIt working group will release BagIt on a open access platform like Github. APTrust, LOCKSS, Fedora Commons, and DSpace will perform the software development 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).
Not applicable.Toolkit, BagIT profile, and software will be created/enhanced with an eye toward sustainability.

B.Workflow and Asset Maintenance/Preservation
B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).
Not applicable. The software products that are being enhanced have their own internal QC processes.

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).
The curation toolkit will be created with sustaining it in mind--the decision to go with textual materials as opposed to a web resource was made while thinking about the migration of the tool going forward.

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).

Not applicable.

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

Not applicable.

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).

Not applicable.

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).

The curation toolkit will be available through the Open Science Framework here: https://osf.io/4d567/. The BagIt profile will be openly available on a platform like Github. The repository and DDP systems make their codebases available on GitHub. All products will have an open source license

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.

Planning grant report is openly available here: doi:10.21985/N28M2Z

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.
The software development that is part of this project is to update/edit existing code of mature products so that they can export bags based on a new shared BagIt profile or accept bags with the new profile. These updates will serve those professionals involved with performing digital preservation actions.

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.
This project is not intending to compete with other software products or create a new product; all the software development is to make existing local repository and distributed digital preservation systems and products more interoperable.

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.
This project involves existing/mature software.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.
Local repository products such as Fedora and DSpace will be updated to export bags based on the new BagIt profile and distributed digital preservation systems, such APTrust will enhance their API so that the new data from the bags can be exposed after ingest into the DDP system. LOCKSS will create a new plugin that supports the profile.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.
Since this project does not intend to create new software, there are no dependencies to mention here.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.
Development will take place over the course of a few months and the project manager will set up weekly check ins where progress updates will be shared. All software products involved in this project have open repositories and extensive publicly available documentation.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.
Northwestern University Libraries: https://github.com/nulib; APTrust: https://github.com/APTrust Fedora Commons: https://github.com/fcrepo4; LOCKSS: https://github.com/lockss

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.

Northwestern University is partnering with multiple open source software projects on this project and is not asserting rights over it. They have different open licenses and the users of the software are well versed in the licenses already since these are mature products. Fedora has the Apache Public license and DSpace, APTrust and LOCKSS are released under the BSD-3 Clause license.

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

The source code for these mature products are available on GitHub. APTrust: https://github.com/APTrust

Fedora Commons: https://github.com/fcrepo4; LOCKSS: https://github.com/lockss

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

Name of publicly accessible source code repository: APTrust: https://github.com/APTrust; Fedora Commons: https://github.com/fcrepo4; LOCKSS: https://github.com/lockss

Part IV: Projects Creating Datasets

Not applicable

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?