# Abstract

DuraSpace seeks a National Digital Platform Planning Grant for $49,279 to investigate barriers upgrading hundreds of U.S.-based libraries and archives running unsupported versions of Fedora. Running unsupported software puts at risk the stability, security, and functionality of the content and services they support.

There are approximately 240 libraries and archives in the United States identified as using unsupported versions of Fedora making them target beneficiaries of the deliverables of this project. They include R1, R2, and R3 universities, liberal arts colleges, and not-for-profit special libraries hosted by historical societies and small research institutes.

DuraSpace's proposed one-year planning project, *Designing a Migration Path: Assessing Barriers Upgrading to Fedora 4.x*, will include consultation with an advisory board of stakeholders, conducting an environmental scan of relevant community initiatives, and gathering primary research data to determine what tools and supports are needed for the upgrade path. We can identify the following as deliverables for this proposed planning project:

- Describing a collection of the most common Fedora 3.x - Fedora 4.x upgrade user stories
- Creating an inventory of tools, documentation, and other resources for the upgrade path
- Providing the Islandora, Samvera, and Fedora community governance bodies feedback on the challenges and advantages of working within their communities
- Developing migration path recommendations and prioritizing recommendations on how to reduce barriers to upgrades that will have the greatest impact on US-based libraries and archives built on Fedora.
- Forming an advisory board to facilitate communication and strategic collaboration among stakeholders that may spur future spin off projects such as:
  - Developing tools to map data models at scale for standard front-end applications
  - Generalizing and documenting existing migration tooling
  - Developing a training curriculum on new standards for description and dependencies for Fedora 4.x
  - Assisting the Islandora community with development of its Fedora 4.x front end (Islandora CLAW)

We believe the direct and indirect deliverables from this project will provide resources and support for the hundreds of American libraries and archives using Fedora to underpin the delivery of scholarly publications, research data, cultural heritage, and special collections-related content and services to patrons.

# Narrative

## 1. Executive Summary

DuraSpace requests a National Digital Platform Planning Grant for $49,279 to investigate barriers to upgrading hundreds of U.S.-based libraries and archives running unsupported versions of Fedora which is putting at risk the stability, security, and functionality of the content and services they support. This project will consult with an advisory board of stakeholders from the Islandora, Samvera, and Fedora communities, conduct an environmental scan of relevant community initiatives, and gather primary research data to inform recommendations to reduce barriers to upgrading to Fedora 4.x. There are approximately 240 U.S.-based libraries and archives identified as target beneficiaries of the deliverables of this project. They include R1, R2, and R3 universities, liberal arts colleges, and not-for-profit special libraries hosted by historical societies and small research institutes.

## 2. Statement of National Need

Hundreds of American libraries and archives use the open source software (OSS) Fedora repository to deliver scholarly publications, research data, cultural heritage, and special collections-related content and services to patrons. Fedora 4.x has been the official release since 2015. However, the upgrade to Fedora 4.x impacts underlying technologies, data models, standards for description, and functionality meaning the upgrade will require re-modelling and migrating data.

The entire Islandora community, which represents 130 repositories in the United States, is still running the previous, now unsupported version of Fedora[1]. Only fifteen percent of these Islandora repositories in the U.S. are hosted by R1 universities[2]. Generally, the Islandora community is characterized by smaller, less resourced institutions and organizations including R2 and R3 universities[3] such as the University of Denver and Andrews University, liberal arts colleges such as Vassar College and Barnard College, library consortia such as the Metropolitan New York Library Council (METRO), and not-for-profit special libraries hosted by historical societies[4] such as the Adventist Digital Library and the California Historical Society.

Similarly, almost 90 custom repositories built on Fedora in the U.S., 48 percent of which are hosted by R1 universities, and 20 Samvera repositories in the U.S., 85 percent of which are hosted by R1 universities, are running unsupported versions of Fedora[56]. Generally, the Fedora and Samvera

---

[1] http://islandora.ca/islandora-installations

[2] A classification indicating Highest Research Activity according to the Carnegie Classification of Institutions of Higher Education

[3] Classifications indicating Higher Research Activity or Moderate Research Activity according to the Carnegie Classification of Institutions of Higher Education

[4] https://docs.google.com/spreadsheets/d/1cKerrKyryvyoM9SU6Uhdw7ROpreJgwlvSzMm9TQ7GTU/edit?usp=sharing

[5] http://registry.duraspace.org/registry/fedora, https://samvera.org/samvera-partners/

communities are characterized by larger, better resourced institutions. Using definitions from Geoffrey Moore's famous Technology Adoption Cycle, Fedora and Samvera implementers are more likely to fall into the innovator or early adopter categories, indicating a willingness to pursue and adopt new technologies, buy into new product concepts very early in their life cycle, and have the insight to match an emerging technology to a strategic opportunity[7]. Members of the Samvera and Fedora communities that are not R1 institutions include Amherst College, Lafayette College, the Rock and Roll Hall of Fame, among others.

Primary research done by DuraSpace in 2017 gathered thirty one stories about major upgrades and migrations. Eight of those stories came from pioneers in the upgrade path from Fedora 3.x to 4.x from the Fedora and Samvera communities.

The subset of Fedora 3.x to 4.x upgrade and migration stories showed the top challenges of the projects were metadata normalization, gaps in skills and knowledge, keeping up with the pace of development in the Fedora and Samvera communities, and redefining services based on the new capabilities of Fedora 4.x. In addition, they shared anecdotes about communications challenges within the Fedora community.

The stories also found the risk of falling out of step with peer and technical support provided by an OSS community was a motivation to upgrade. For example, one respondent said "We didn't want to be stuck on abandoned software. It was more about sustainability. [...] We don't have the resources to build a custom solution."[8] Unfortunately the majority of Fedora implementers find themselves in this situation. At least 240 U.S.-based libraries and archives are running unsupported versions of Fedora. Running unsupported versions of software carries risks including, "[losing the] stability of a mainstream code release, the risk to information security, and the likelihood that the tool in question will become increasingly less functional and reliable as it ages."[9]

Unfortunately, general and reusable resources for the upgrade path are lacking. The pioneers of the Fedora 3.x to 4.x migrations noted access to peers via Slack, Google Groups, and chat rooms (IRC), documentation and existing code or scripts as the primary resources leveraged in their projects. Coordination of the transfer and review of these artifacts is required to begin forming a common and complete tool set for staff and service providers working for smaller and less resourced organizations, to attempt the migration path.

The question of how many staff resources are dedicated to digital collections management was investigated as part of the Bridge2Hyku (LG-70-17-0217-17) project. In a project update provided in May 2018, the University of Houston (UH) Libraries shared results of a survey of four partner institutions for the project. The survey found that an average of "7.625 FTE support digital collections

---

[6]https://docs.google.com/spreadsheets/d/1cKerrKyryvyoM9SU6Uhdw7ROpreJgwlvSzMm9TQ7GTU/edit?usp=sharing

[7] Moore, Geoffrey, 1999, Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers, Harper Business. Pages 13, 256

[8] Tripp, E. (2018, Jan 21). Anonymized Migration Stories Survey Results 2018. Retrieved from https://osf.io/36pmc/

[9] Gengenbach, M., Peltzman, S., Meister, S., Graham, B., Waugh, D., Moran, J., Seifert, J., Dowding, H., and Carleton, J. (2016, Oct 25). OSS4EVA: Using Open-Source Tools to Fulfill Digital Preservation Requirements. Retrieved from http://journal.code4lib.org/articles/11940

management, roughly 6%. When asked how many FTE of local IT staff time is devoted to digital collection management, the average drops to 2%".[10] In a personal communication with Andrew Weidner, the Bridge2Hyku Project Manager, he said "We were surprised by how little staff support digital collections receive among our partner institutions. If that pattern holds true for the broader community, that could explain some of the difficulties that others are having with migrating to Fedora 4.x."

In contrast to the amount of staff resources dedicated to digital collections management, members of the Fedora, Islandora, and Samvera communities are using Fedora 3.x to provide access and preservation for a wide array resources, underpinning services for patrons. Anecdotes from our advisory board tells us that a single repository can store millions of objects and hundreds of terabytes (TB) of unique content in formats including books, manuscripts, maps, photographs, oral histories, music, video, data, web sites, theses, dissertations, journal articles, 3D objects, and so forth that are resources for faculty, researchers, and students[11]. Other advisory board members tell us that their Fedora 3.x repositories have an emphasis on digital preservation and library-managed at-risk collections. The intellectual output of the campus and exceptional collections of scholarly and educational research materials are stored in Fedora 3.x and are part of faculty resolutions for supporting open access. Advisory board members tell us they take seriously the role of representing the digital repository needs of smaller institutions, that work closely with peers through partnerships to achieve shared goals. We feel these anecdotes to be representative of members of the broader Fedora and Fedora-based communities, the majority of which are libraries and archives hosting vital patron services on unsupported software.

We feel strategic collaboration and sharing resources will be fundamental in creating the support needed for migrations to Fedora 4.x and mitigating the risks of running patron services on unsupported software. Our recent primary research on migrations and upgrades indicated that facilitating communication among stakeholders will be a big part of the job. One of the interviewees advised,

> "A migration is related to a service and program upgrade. It involves trust and credibility. Relationship building is necessary so you can address concerns with change and maintain trust. Moving from F[edora] 3[.x] to F[edora]4[.x] has been spoken about lot in the Fedora community since 2012. It wasn't about technology. The issue is 'how do I maintain my data and protect from failure? [...] We need to connect our communities so we share our knowledge. Projects are 80% communications and 20% technology. [...].

DuraSpace staff and members of the advisory board agree that more communication is needed and can be achieved through this proposed project. We see an immediate need for consultation and communications support as it will work to resolve existing challenges. We can point to two communications scenarios encountered in 2017 that illustrate this need:

1) A pioneer of the Fedora 3.x to Fedora 4.x migration path who was interviewed in the DuraSpace

---

[10] Crocken, T. (2018, May) Initial Report from the Bridge2Hyku Digital Collections Survey. Retrieved from https://docs.google.com/document/d/1z6loFo66fzz7uBBm9g8B6m03fudG0LdgHLAtF72X9XQ/edit#heading=h.5wz6bqd6 qv5b

[11] https://library.stanford.edu/research/stanford-digital-repository/sdr-overview

migration stories survey experienced difficulty with two of Fedora 4.x's dependencies for their particular use cases. The interviewee said the migration "failed because at the time we were the largest repository to try to migrate. Many others started green field[12]. Ours turned out to be disastrous. There were problems with the underlying Infinispan and ModeShape components when dealing with large files." The same project resulted in a community discussion about how Fedora 4.x provides preservation support. The interviewee said "When we move to F[edora]4.x we'll have to totally change the way we do preservation." This feedback was shared and discussed widely in the Fedora community and led to a survey to gauge preservation expectations and capacity[13], the Fedora API specification project, and the Oxford Common Filesystem Layout (OCFL)[14] project. Despite these efforts, Fedora 4.x lost important momentum and requires renewed effort to communicate the outcomes of these initiatives as well as long-term strategic collaborations to stay abreast of requirements and collect feedback.

2) The Islandora community is working on the development of its next generation release with Fedora 4.x as the backend, called Islandora CLAW. The idea of describing content with RDF instead of XML-based metadata standards was initially met with resistance at the Islandora Conference in May 2017. The unfortunate term "MODSpocalypse" was dubbed and used to highlight the unease expressed over this change and the current gap in skills and knowledge within the community related to RDF. Leaders in the community have been working diligently to change the conversation to focus on implementing RDF and linked data as a top technology trend[15] that will enhance discovery and interoperability, dubbing the transition as the "RDFaissance". This also points to the need for prioritizing the development of resources and support for stakeholders anticipating the Fedora 3.x to Fedora 4.x migration path.

DuraSpace's proposed planning project, *Designing a Migration Path: Assessing Barriers Upgrading to Fedora 4.x*, will include consultation with an advisory board of stakeholders from the Islandora, Samvera, and Fedora communities, conducting an environmental scan of relevant community initiatives, and gather primary research data to inform recommendations for what other tools and supports are needed to support majority adopters in the Fedora, Samvera and Islandora communities. It will also review migration tools and resources created by the Fedora 3.x to 4.x. upgrade pioneers to determine what can be generalized, tested, and documented for others to use. The cross-community engagement and information gathering proposed in this project will lead to spin off projects that will define, plan and allocate resources and support for a migration path to Fedora 4.x.

Our work will complement related projects including Bridge2Hyku (LG-70-17-0217-17) and Beyond the Repository (LG-72-16-0135-16), on which DuraSpace staff are advisory board members. We have received letters of support from representatives of these projects, among others from the Samvera Community, Islandora Foundation, and the OCFL Editorial Committee.

---

[12] Meaning from scratch without the need for migration.

[13] http://www.duraspace.org/fedora/resources/publications/fedora-digital-preservation-survey/

[14] https://docs.google.com/document/d/13gFfSu1fePKx0eQYk458zE6OZwCW5JH_hTIG58f3S8c/edit

[15] http://lj.libraryjournal.com/2015/02/shows-events/ala/lita-members-talk-tech-trends-ala-midwinter-2015/#_

## 3. Project Design

The project has a proposed timeline of one year, running from Oct 1, 2018 to Sept 30, 2019. However, we feel the deliverables can be achieved in nine months. The project is organized into three phases, including 1) consultation, 2) information gathering and 3) evaluation and dissemination. Tasks and work effort estimates for this project were developed using an iterative group process to challenge assumptions and clarify workflow called the Wideband Delphi estimation technique[16]. Estimates are also based on our experience conducting similar types of projects. The work effort for DuraSpace staff is estimated at 653 hours.

Personnel for this project include Andrew Woods, Fedora Technical Lead, David Wilcox, Fedora Product Manager, Daniel Bernstein, DuraSpace Technical Lead and Developer, and Erin Tripp, DuraSpace Business Development Manager and Interim CEO. Both Mr. Woods and Mr. Bernstein will provide technical expertise. Mr. Woods, Mr. Wilcox, and Mr Bernstein will provide knowledge of related initiatives in the community and connections to stakeholders. Ms. Tripp will provide expertise in conducting research and project design.

Members of the advisory board include:

- Andrew Weidner, Digital Operations Coordinator & Bridge2Hyku Project Manager, University of Houston Libraries
- Este Pope, Head of Digital Programs, Amherst College
- Mark Jordan, Chair of Islandora Foundation Board of Directors & Head of Library Systems, Simon Fraser University
- Mike Giarlo, Member of Samvera Steering Committee & Software Engineer & Architect, Stanford University
- Sayeed Choudhury, Associate Dean for Research Data Management, Johns Hopkins University
- Tim Shearer, Associate University Librarian for Digital Strategies and IT, University of North Carolina at Chapel Hill

The Advisory board members represent a spectrum of experiences deploying Fedora, as well as, a diversity of organization sizes and resources including community governance representatives, R1 institutions, a representative of a related grant-funded project, and a liberal arts college that collaborates with consortia-style bodies for sharing resources and reducing costs.

**Phase One: Consultation (Oct 2018 - Dec 2018)**

**Task 1:** Establish and meet with the advisory board to kick off project (Oct 2018)
**Description:** A virtual meeting for introductions and to communicate the schedule, goals, deliverables of the project as well as discuss methodology for the survey and environmental scan.
**Estimate:** Ten (10) hours of work effort for three DuraSpace staff to prepare for and run the meeting.

---

[16]https://docs.google.com/spreadsheets/d/1jCX39orARuLCKh1miqXP2YVGfF-SKOWy0V7cXkOl_YE/edit?usp=sharing

**Task 2:** Conduct an environmental scan of literature on software upgrades and migrations as well as planned or recommended Fedora 3.x - Fedora 4.x upgrade projects (Oct 2018)
**Description:** Review previously conducted surveys and recent publications about Fedora 3.x - Fedora 4.x migrations. The advisory board can help recommend information sources.
**Estimate:** 28 hours of work effort for two DuraSpace staff to search for and review relevant literature. One DuraSpace staff member will ask for recommended sources of information from advisory board and summarize the information.

**Task 3:** Develop First Draft Survey (Nov 2018)
**Description:** Targeted at Fedora 3.x users, it will be designed to determine the reasons for not moving to Fedora 4.x and how to reduce barriers for the upgrade.
**Estimate:** 22 hours of work effort for two DuraSpace staff to provide input on survey questions and for one DuraSpace staff member to create the survey and send it to the advisory board to collect feedback and iterate as needed.

**Task 4:** Subset Consultations with the Samvera, Islandora and Fedora users  (Nov - Dec 2018)
**Description:** Virtual communications to iterate on survey design, challenge assumptions, and gauge interest in Fedora 4.x, its capabilities, and barriers to upgrading. Conducting survey interviews with each member of the advisory board individually and one or two more people, 6-7 total. Expecting at least one hour per interview.
**Estimate:** 30 hours of work effort. Two DuraSpace staff will iterate on survey design, 8 hours each. One DuraSpace staff member will conduct survey interviews in 14 hours.

**Task 5:** In-Person Advisory Board meeting at CNI in Washington (Dec 2018)
**Description:** In-person meeting to review environmental scan, survey, and solicit advisory board for contacts to include in survey distribution.
**Estimate:** 90 hours of work effort for three DuraSpace staff to prepare for the meeting, attend the full day meeting in-person, and travel time (30 hours each).


**Phase Two:  Information Gathering (Jan 2019 - Mar 2019)**


**Task 1:** Review of the front-end applications used to interact with Fedora 3.x in order to determine commonalities/differences (Jan 2019)
**Description:** Assess commonalities among application specifications and impact to migration/export to Fedora 4.x through review of documentation of general and custom implementations (8 in total). We will review Islandora and Samvera specifications that are available. Many custom sites are not specified and documented. We aim to include reviews of different types of implementations. The

criteria considered will include implementers of the Portland Common Data Model (PCDM), examples of small and large repositories, examples of repositories at small and large institutions, and repositories catered for institutional repository, data management, images and manuscripts and other archival use cases. We will ask the advisory board for leads in our selection process.

**Estimate:** 36 hours of work effort total. Two DuraSpace staff (16 hours each) to review front end application specifications and provide a summary to the advisory board. 4 hours for one DuraSpace staff member to review the summary, send to the advisory board, and gather feedback.

**Task 2:** Collect and review Fedora 3.x data from stakeholder groups (Jan 2019)
**Description:** Assess commonalities among approaches to data modelling and impact to migration/export to Fedora 4.x through data review. We will review data from the 8 implementations reviewed in phase 2, task 1. We will request a 10% representative sample set, including edge cases, and we may put a limit on the size of the set of data to review.
**Estimate:** 34 hours of work effort total. Two DuraSpace staff (16 hours each) to review data from the implementations selected and provide a summary (including instances of when documentation does not match with data model/implementation) to the advisory board. 2 hours for one DuraSpace staff member to review the summary, send to the advisory board, and gather feedback.

**Task 3:** Review existing migration/import/exporting tooling (Jan 2019)
**Description:** Determine applicability of tooling for Fedora front-end applications. Migration tooling isn't generalized for community use. We know of two sets of tools at the moment.
**Estimate:** 12 hours of work effort, including 8 hours for one DuraSpace staff member to collect and summarize information and 4 hours for one DuraSpace staff member to review and send to the advisory board, and collect feedback.

**Task 4:** Review outcomes of the Fedora API specification project (Jan 2019)
**Description:** Summarize impacts of the new Fedora API specification on migrations/upgrades. We assume the biggest impact will be on the client side, not the data migration side. It could provide opportunities for users to implement variant backend applications.
**Estimate:** 18 hours of work effort including 8 hours each for two DuraSpace staff to collect, discuss impacts, and summarize information. Includes 2 hour for one DuraSpace staff member to review and send summary to the advisory board, and collect feedback.

**Task 5:** Review ongoing work on Oxford Common Filesystem Layout (OCFL) (Jan 2019)
**Description:** Summarize impacts of OCFL on migrations/upgrades.
**Estimate:** 5 hours of work effort including 2 hours each for two DuraSpace staff to summarize information. One hour for a DuraSpace staff member to gather feedback from the advisory board.

**Task 6:** Finalize Survey (Jan 2019)

**Description:** Iterating on the survey based on the information gathered, reviewed, and discussed.
**Estimate:** 5 hours of work effort for a DuraSpace staff member to reflect on information gathered, make changes, and ask for feedback from team and advisory board and iterate again as needed.

**Task 7:** Administer Survey to stakeholder community (Feb-Mar 2019)
**Description:** Distributing the survey and the project team undertaking one on one interviews using the survey. Aiming for 80 responses from various staff positions including managers, librarians/archivists, technologists, etc. We will distribute the survey to lists including Samvera-Tech, Samvera-Community, Digital-Curation, Code4Lib, DLF, PASIG, Fedora-Community, Islandora-Community, Islandora-Dev. In addition, we will send personal invitations to complete the survey online or as an interview to individual stakeholders we would like to participate. In our experience, personal invitations from project staff and advisory board members have a high success rate.
**Estimate:** 38 hours of work effort including 30 hours for a DuraSpace staff member to prepare messaging, invitation recipient lists, make assignments to members of the team and advisory board to issue invitations and deliver a webinar on the topic. Two DuraSpace staff are allocated 4 hours each to undertake assignments and participate.

## Phase Three: Evaluation and Dissemination (Apr 2019 - Jun 2019)

**Task 1:** Anonymize survey of all personally identifying information (PII) and identify gaps in the results (April 2019)
**Description:** Anonymize data for distribution and identify gaps in the results.
**Estimate:** 36 hours including 22 hours for a DuraSpace staff member to anonymize and identify gaps in the results to inform follow up questions for secondary consultation. 14 hours for two DuraSpace staff (7 each) to conduct detailed review of 80+ responses to confirm removal of all PII.

**Task 2:** Secondary consultation with Samvera, Islandora and Fedora Communities (April 2019)
**Description:** Conduct 10 in depth interviews with survey respondents via virtual communications to fill in gaps in data set. Conducting survey interviews with each member of the advisory board individually and 3-4 more people.
**Estimate:** 14 hours of work effort including 10 hours for a DuraSpace staff member to conduct interviews and 4 hours to summarize and send results to the advisory board for review.

**Task 3:** In-Person Advisory Board meeting at CNI in St. Louis (April 2019)
**Description:** Roundtable discussion at CNI Spring meeting 2019 to share, react to, and discuss survey results and gather feedback from the advisory board.
**Estimate:** 90 hours of work effort for three DuraSpace Staff for meeting prep, attending a full day

meeting in-person, and travel time (30 hours each).

**Task 4:** Write the final report and recommendations (May-June 2019)
**Description:** The report will summarize all information gathered including a collection of the most common Fedora 3.x - Fedora 4.x upgrade user stories, provide an inventory of tools, documentation and other resources that already exist, feedback and anecdotes to Fedora, Samvera, and Islandora community governance bodies of challenges and advantages of working within the communities, and prioritized recommendations on how to reduce barriers to upgrades that will have the greatest impact on U.S.-based libraries and archives built on Fedora.
**Estimate:** 88 hours of work effort including 40 hours for a DuraSpace staff member to compile and write report and send to team/advisory board for feedback. Estimate includes time to debate and discuss recommendations and priorities relative to identified pain points. Estimate also includes 24 hours for two DuraSpace staff to consult on user stories, review materials, and provide feedback.

**Task 5:** Disseminate final report and recommendations (June 2019)
**Description:** DuraSpace has a broad communications network including more than 3,500 followers on the @Duraspace and @FedoreaRepo twitter accounts and more than 6,000 subscribers to the DuraSpace Digest, Member Newsletter, and Fedora Newsletter. We intend to publish the report and recommendations from this project on our wiki, website, publish a blog post, and disseminate direct links to our communications network. In addition, we will create a base proposal and slide deck template that can be used for speaking engagements about the project at conferences DuraSpace staff and advisory board members usually attend such as Fedora User Group Meetings and Camps, Open Repositories, DLF, Code4Lib, and CNI. The cost for DuraSpace staff to travel to these events is already included in DuraSpace annual operating budget. Therefore, we are estimating the workeffort to publish, disseminate via our social media and newsletter assets, as well as, write a base presentation proposal that can be modified as needed and submitted for event talks.
**Estimate:** 12 hours of work effort including publishing and posting to our communications network and preparing a base presentation proposal and slide deck that can be re-used and modified by speakers. This estimate does not include travel time or presentation delivery time.

# 4. National Impact

Bridging the gap between unsupported versions of Fedora and the current, supported release, and improving consultation and communications practices in the community are essential to safeguarding the digital heritage entrusted to the Fedora community. This project will pave the way for a spectrum of organizations to upgrade to Fedora 4.x more easily with fewer resources.

We can identify the following as direct deliverables for this proposed planning project:

- Describing a collection of the most common Fedora 3.x - Fedora 4.x upgrade user stories

- Creating an inventory of tools, documentation, and other resources for the upgrade path
- Providing the Fedora, Samvera, and Islandora community governance bodies feedback and anecdotes of challenges and advantages of working within their communities
- Forming an advisory group to facilitate communication and long-term strategic collaboration among stakeholders that may facilitate future spin off projects
- Developing migration path recommendations and prioritizing recommendations on how to reduce barriers to upgrades that will have the greatest impact on US-based libraries and archives built on Fedora. We will format the materials so they are usable, discoverable, reusable, and have the potential to be generalized for use in other OSS communities such as DSpace

These deliverables will have a direct impact on the organizations that are planning to execute a Fedora 3.x - Fedora 4.x upgrade. The deliverables have the potential to:

- Spur increased engagement, communications, transparency, trust within the Fedora, Samvera, and Islandora communities
- Provide information resources for a migration business case to institutional decision makers
- Form the basis for migration project planning and budgeting
- Encourage the allocation of staff and knowledge sharing to fill gaps in community migration tooling and other resources
- Build momentum for Fedora 4.x upgrades and development for other related initiatives such as the Fedora API specification project, OCFL, and Bridge2Hyku

We identify the following as possible spin-off initiatives for this project:

- Developing tools to map data models at scale for standard front-end applications
- Generalizing and documenting existing migration tooling
- Developing a training curriculum on new standards for description and dependencies for the Fedora community that has the potential to be shared in other OSS communities as well
- Assisting the Islandora community with development of its Fedora 4.x front end (Islandora CLAW)

We believe the direct and indirect deliverables from this project will increase the number of U.S.-based libraries and archives using supported software to safeguard and deliver content and services to patrons. Also, creating a critical mass of Fedora 4.x implementations will advance digital preservation theory and practice because Fedora 4.x's linked data implementation enhances discovery[17] and supports new, more sustainable models for extensibility and integration with external applications. It will further facilitate technical development and collaboration around community supported, free, and open source repository software.

---

[17]http://lj.libraryjournal.com/2015/02/shows-events/ala/lita-members-talk-tech-trends-ala-midwinter-2015/#_

| | October 2018 | November 2018 | December 2018 | January 2019 | February 2019 | March 2019 | April 2019 | May 2019 | June 2019 |
|---|---|---|---|---|---|---|---|---|---|
| **Phase One: Consultation** | ▓ | ▓ | ▓ | | | | | | |
| Task 1: Establish and meet with the advisory board to kick off project | ■ | | | | | | | | |
| Task 2: Conduct an environmental scan of literature on software upgrades and migrations as well as planned or recommended Fedora 3.x - Fedora 4.x upgrade projects | ■ | | | | | | | | |
| Task 3: Develop First Draft Survey | | ■ | | | | | | | |
| Task 4: Subset Consultations with the Samvera, Islandora and Fedora users (Nov - Dec 2018) | | | ■ | | | | | | |
| Task 5: In Person Advisory Board meeting at CNI in Washington | | | ■ | | | | | | |
| **Phase Two: Information Gathering** | | | | ▓ | ▓ | ▓ | | | |
| Task 1: Review of the front-end applications used to interact with Fedora 3.x in order to determine commonalities/differences | | | | ■ | | | | | |
| Task 2: Collect and review Fedora 3.x data from stakeholder groups | | | | ■ | | | | | |
| Task 3: Review existing migration/import/exporting tooling | | | | ■ | | | | | |
| Task 4: Review outcomes of the Fedora API specification project | | | | ■ | | | | | |
| Task 5: Review ongoing work on Oxford Common Filesystem Layout (OCFL) | | | | ■ | | | | | |
| Task 6: Finalize Survey | | | | | ■ | | | | |
| Task 7: Administer Survey to stakeholder community | | | | | ■ | ■ | | | |
| **Phase Three: Evaluation and Dissemination** | | | | | | | ▓ | ▓ | ▓ |
| Task 1: Anonymize survey of all personally identifying information (PII) and identify gaps in the results | | | | | | | ■ | | |
| Task 2: Secondary consultation with Samvera, Islandora and Fedora Communities | | | | | | | ■ | | |
| Task 3: In Person Advisory Board meeting at CNI in St. Louis | | | | | | | ■ | | |
| Task 4: Write the final report and recommendations | | | | | | | | ■ | ■ |
| Task 5: Disseminate final report and recommendations | | | | | | | | | ■ |

# 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 digital content, resources, and assets created as deliverables and work products of our proposed project will be under the copyright of Fedora Commons,d.b.a, DuraSpace and released to the public using a CC BY 4.0 license. We use this license for all of our wiki content and documentation.

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.

> We will assert no additional ownership rights over the deliverables and work products other than those included in the CC BY 4.0 license.

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.

> We will gather express consent from the advisory board so that meeting minutes can be recorded on the DuraSpace wiki, providing public access.

> In the creation of the survey, we will include a preamble such as "The respondent is participating in an IMLS National Digital Platform Planning Project that will lead to an anonymous set of survey results made available to the public, a recommendations report, and presentation. The

respondent is free to discontinue the interview/survey at any time. Respondent names will be collected for administrative purposes only and will not appear in any published work unless the permission of the respondent has been obtained in writing." Also, we have allocated time in our estimates for this project to anonymize survey data before it is made available to the public.

## 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.

Our proposed project will produce:

- One set of anonymized survey results in Google Sheets format with with permissions to view and comment, as well as .PDF and .CSV formats.
- One recommendations report and multiple appendices that will be made available in Google Docs with permissions to view and comment, as well as .PDF format.
    - Appendices will include summaries of the 1) environmental scan of relevant community initiatives, 2) review of documented Fedora front-end applications and 3) related data, 4) review of existing migration/import/exporting tooling, 5) review progress on the Fedora API Spec and 6) Oxford Common Filesystem Layout (OCFL), and 7) anonymous survey results.
- One template presentation proposal that will be made available in Google Docs format with permissions to view, copy, edit, and comment, as well as .PDF format.
- One template presentation (including speaking notes) that will be made available in Google Slides format with permissions to view, copy, edit, and comment, as well as .PDF format and other formats as requested.

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.

DuraSpace uses the Google suite of applications including Gmail, Docs, Sheets, Slides, and Forms for communications, word processing, etc. DuraSpace will undertake the work related to creating deliverables and work products for this proposed project.

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).

We plan to use the following formats: Google Forms, Docs, Sheets, Slides, as well as .PDF, and .CSV. We can provide other formats on request such as .odt, .txt, .docx, or pptx.

### 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).

All DuraSpace personnel and advisory board members listed on the List of Key Project Staff and Consultants will have read and write permissions to the digital assets in Google Docs and Sheets, with the exception of the survey results. Read and write permissions for the original survey results will be limited to Ms. Tripp as the asset will include personally identifying information (PII). The anonymized set will be shared with the team and advisory board.

Google Docs and Sheets tracks versions of the assets, including user and timestamps of edits

that will assist team oversight of asset quality control and workflow.

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).

> Final copies of the assets can be uploaded to the DuraSpace/Fedora wiki and websites and deposited into one of our advisory board member institution's repositories for discovery, as well as uploaded in the DuraSpace DuraCloud account for duplicated preservation storage. We will include these tasks in the project plan.

## 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).

> We will provide descriptive metadata for the assets deposited into one of our advisory board member institution's repositories. The descriptive schema used will comply with the requirements of submission (likely DC or MODs). As a Fedora-based repository, it will automatically generate technical and administrative metadata. If an Islandora repository is selected (e.g. advisory board member Mark Jordan's repository at Simon Fraser University) PREMIS data will also be generated. However, the selection of repository will be made at a later date. We will also provide basic descriptive metadata for the DuraCloud upload. DuraCloud will automatically generate technical metadata. There is no native metadata schema or standard in DuraCloud.

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

> The preservation and maintenance of the metadata and files will be facilitated via repository management at the institution where the assets are deposited as well as through the DuraCloud upload. DuraCloud content is automatically duplicated and integrity checked at least bi-annually.

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).

> We will present the findings on this proposed project in conference/event presentations and a webinar. In those presentations we will point to the locations of the digital assets by weblink. Islandora and Samvera repositories can be harvested via OAI-PMH and semantic web protocols. In addition, the DuraSpace/Fedora wiki and website is crawled by search engines so assets shared there will be discoverable via Google, Bing, etc.

## 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 digital assets will be made openly available online through our wiki and website and by

depositing them into an open access repository (Fedora, Islandora, or Samvera) to be selected at a later date. An example of hardware/software platforms that would be used is available in the description of the Stanford Digital Repository (SDR). They describe leveraging Fedora, Blacklight, and Solr as applications and spinning disk and LTO storage with multiple copies in diverse locations. The files will be made available in non-proprietary formats such as .PDF and .CSV that will not require fee-based software to read.

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.

We submit the Digital Preservation Survey report and survey results as examples:

- https://duraspace.org/fedora/resources/publications/fedora-digital-preservation-survey/
- https://duraspace.org/fedora/resources/publications/fedora-digital-preservation/
- https://duraspace.org/fedora/resources/publications/fedora-digital-preservation-survey/#downloads

# 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.

Not applicable for this project.

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

Not applicable for this project.

## B. Technical Information

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

Not applicable for this project.

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

Not applicable for this project.

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

Not applicable for this project.

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

Not applicable for this project.

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

Not applicable for this project.

## C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under

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

> Not applicable for this project.

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

> Not applicable for this project.

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

Name of publicly accessible source code repository:

URL:

> Not applicable for this project.


## 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.

> Not applicable for this project.

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?

> Not applicable for this project.

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).

> Not applicable for this project.

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.

> Not applicable for this project.

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).

> Not applicable for this project.

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?

> Not applicable for this project.

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

> Not applicable for this project.

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

Name of repository:

URL:

> Not applicable for this project.

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

> Not applicable for this project.