Fedora 4 API Extension (API-X) Architecture

The Sheridan Libraries at Johns Hopkins University request $120,500 from the Institute of Museum and Library Services (IMLS) to develop a service proxy layer on top of the Fedora 4 software platform that will facilitate the exposure of repository contents and services as linked data web resources. This work was initially inspired by a redesign and refactoring of the Data Conservancy[1] software, which was initially prototyped by the Sheridan Libraries and a network of partners through a grant from the National Science Foundation's (NSF) DataNet program. Following the end of DataNet funding, the Data Conservancy decided to apply the lessons learned and redesign and refactor the entire architecture and software platform. The re-imagined Data Conservancy infrastructure builds upon Fedora 4, a digital object repository platform whose development is led by the nonprofit DuraSpace[2] organization.

IMLS wisely emphasizes that any national digital platform must leverage and extend existing technology. There is an existing user base of Fedora software for institutional repositories that will grow given the important enhancements and robustness offered by Fedora 4. By providing architecture to deploy repository services as lightweight extensions, institutions that use Fedora 4 for their institutional repository needs would be automatically positioned to extend their platforms for more robust data management. As federal funding agencies respond to the White OSTP memoranda regarding public access to publications and data, it is becoming clear that simply depositing and subsequently downloading data will not be sufficient. The proposed work supports a vision of data management where data are packaged with information graphs that capture and preserve connections to publications and software.[3]

The direct audience for the proposed API-X work is Fedora 4 software developers and repository managers who support a range of cultural heritage institutions including libraries, museums and archives. The Fedora user community has been growing steadily through adoption of Fedora 3 and a set of associated applications such as Hydra, Islandora and Avalon. Fedora 4 represents a major advance particularly related to linked data capabilities. The API-X work represents an opportunity to augment Fedora 4 by exposing services on repository resources in a manner consistent with REST and linked data best practices.

The proposal describes software design, software development and community engagement activities that would occur from March 2016 through December 2016 with IMLS funding requested starting May 2016. The software development will be evaluated based on the agile methodology that focuses fixed period (typically two weeks) sprints based on a points system for measuring performance. The overall impact of API-X will be measured based on adoption of the extensions described within this proposal and the development of new extensions based on the API-X framework.

---

[1] http://dataconservancy.org

[2] http://duraspace.org

[3] For example: https://www.youtube.com/watch?v=YsZLUdJDFX0

Fedora 4 API Extension (API-X) Architecture

1. Statement of Need

The Sheridan Libraries at Johns Hopkins University request $120,500 from the Institute of Museum and Library Services (IMLS) to develop a service proxy layer on top of the Fedora 4 software platform that will facilitate the exposure of repository contents and services as linked data web resources. This work was initially inspired by a redesign and refactoring of the Data Conservancy[1] software, which was initially prototyped by the Sheridan Libraries and a network of partners through a grant from the National Science Foundation's (NSF) DataNet program. Following the end of DataNet funding, the Data Conservancy decided to apply the lessons learned and redesign and refactor the entire architecture and software platform. The re-imagined Data Conservancy infrastructure builds upon Fedora 4, a digital object repository platform whose development is led by the nonprofit DuraSpace[2] organization.

While Fedora 4 offers a REST API for basic read/write operations, it does not provide any means to extend this API or expose services on repository resources. The Data Conservancy software team came to the conclusion that if such a service and extension framework existed, much of the Data Conservancy software and other software platforms could be realized as a series of lightweight add-ons to existing Fedora repositories. The Data Conservancy team presented the idea of such a framework to Product Manager and Technical Lead for Fedora and to the community during the Fedora User Group meeting at the 2015 Open Repositories Conference. As a result of these community engagements, this proposed Fedora 4 API extension concept ("API-X") has been adopted as a community project managed under the auspices of DuraSpace.[3]

The first step in this community development process is the collection of use cases. As noted from a link[4] from the API-X wiki space, multiple institutions have submitted thirty use cases as of January 2016. It is worth noting that while the original drivers for API-X relate to the linked data and preservation-oriented approach to data management identified by the Data Conservancy, the range of use cases and interested institutions provides evidence that this proposed work has broader utility and relevance.

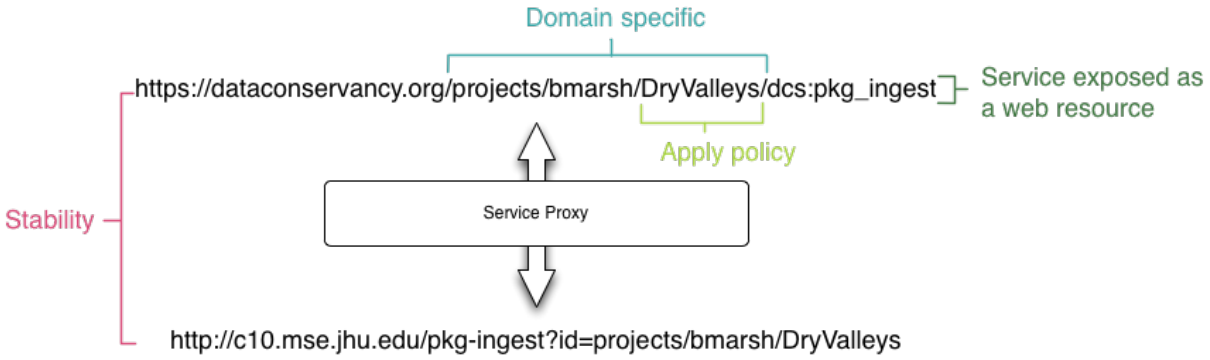
IMLS wisely emphasizes that any national digital platform must leverage and extend existing technology. There is an existing user base of Fedora software for institutional repositories that will grow given the important enhancements and robustness offered by Fedora 4. By providing architecture to deploy repository services as lightweight extensions, institutions that use Fedora 4 for their institutional repository needs would be automatically positioned to extend their platforms for more robust data management. As federal funding agencies respond to the White OSTP memoranda regarding public access to publications and data, it is becoming clear that

[1] http://dataconservancy.org
[2] http://duraspace.org
[3] https://wiki.duraspace.org/display/FF/Design+-+API+Extension+Architecture
[4] https://wiki.duraspace.org/display/FF/Use+Cases+-+API+Extension+Architecture

simply depositing and subsequently downloading data will not be sufficient. The proposed work supports a vision of data management where data are packaged with information graphs that capture and preserve connections to publications and software.[5]

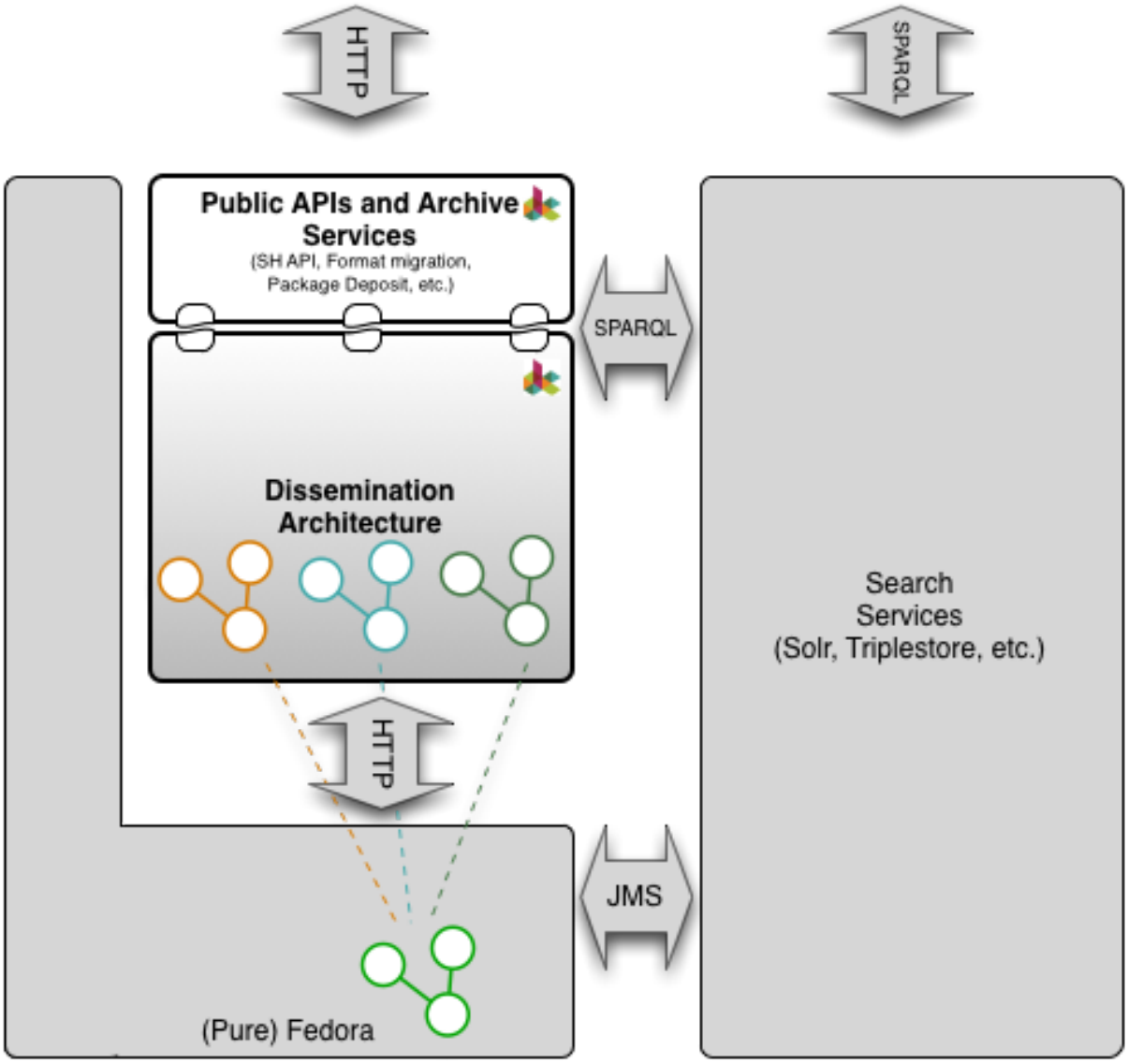
The direct audience for the proposed API-X work is Fedora 4 software developers and repository managers who support a range of cultural heritage institutions including libraries, museums and archives. The Fedora user community has been growing steadily through adoption of Fedora 3 and a set of associated applications such as Hydra, Islandora and Avalon. Fedora 4 represents a major advance particularly related to linked data capabilities. The API-X work represents an opportunity to augment Fedora 4 by exposing services on repository resources in a manner consistent with REST and linked data best practices.

Repository middleware may be deployed across a variety of backend hosts or servlet containers, and may even be replicated/clustered for performance reasons. This deployment is considered an implementation detail that should not have to be known by users of the underlying archive or its associated services. The view that is proposed via API-X to the public is a coherent, stable hierarchy of cool URIs[6] exposing resources that link to one another as depicted below:



The resources exposed by the repository are intended to follow linked data principles and be publishable on the web as linked data. This means that resources *should* refer to one another by URIs that can be followed by a user agent (i.e., dereferencable). Furthermore, these URIs *should* be cool URIs that are stable over time, point to resources that may have different representations retrievable by content negotiation, and do not conflate representation and description. The same is true of the resources and URIs exposed by services or middleware over the contents of the repository that constitute a public API. To this end, the Sheridan Libraries is proposing that API-X serve as a sort of HTTP service proxy, routing requests public URIs to the services that implement them as depicted below:

[5] For example: https://www.youtube.com/watch?v=YsZLUdJDFX0
[6] http://www.w3.org/TR/cooluris/



Within this overall architecture, the core Fedora software itself would not need to be enhanced or modified in any way. Its role would largely be limited to:

- Publishing the URIs of repository resources.
- Identifying the rdf:types of a given Fedora object, when necessary.
- Allowing backend services to CRUD (create, replace, update, delete) Fedora objects as necessary.

API-X is envisioned as the service proxy layer that may be implemented by infrastructure such as Apache Camel, deployed separately from Fedora itself. This differs from the notion of

'disseminators' in the Fedora 3 sense, where Fedora itself acted as a very specialized kind of proxy that required knowledge of how to invoke individual web services as defined by WSDL and parameter bindings for each dissemination. The concept of 'disseminators' in Fedora 3 was not carried forward to Fedora 4 due to this complexity. In contrast, API-X in the Fedora 4 sense may involve simple URL rewriting and an HTTP reverse proxy, forwarding entire http requests to the services that consume them.

2. Impact

Fundamentally, the proposed extensions to Fedora 4 APIs will result in resources within Fedora 4 repositories being exposed as linked data resources and facilitate use of these resources through services within the entire lifecycle of content management. By moving the Data Conservancy software to a series of lightweight add-ons to Fedora 4 repositories, libraries, museums and archives would be able to use the core software platform as an institutional and data repository in a fully robust, semantic manner. Rather than assume data management constitutes deposit and subsequent download of data, the proposed API-X work would move data management into the realm of automatically understanding connections between publications, data and software and the available services to act upon them. This latter vision of data management more readily supports the ultimate goals such as reproducibility of research and long-term preservation.

On a more specific level, multiple software developers and repository managers have submitted the thirty use cases and participated in a consensus-building process of requirements definition and prioritization. The use case development, collection and analysis have been conducted under the auspices of DuraSpace, which acts as a clearinghouse and facilitator for community-based Fedora software development.

The use cases describe an array of capabilities ranging from content ingestion (e.g., deposit status and update, recovering from interrupted deposit), storage management (e.g., compound objects in bags stored in cloud-based service), digital preservation (e.g., file derivatives, format migration), data transformation (e.g., render Fedora objects as MODS/XML), access management (e.g., embargo), content characterization (e.g., graph-based representation), content presentation (e.g., image rendering) and integration with discovery protocols (e.g., OpenSearch). In addition to this diverse array of capabilities, the proposed API-X would support novel approaches for interoperability of linked data resources such as signposting (Van de Sompel and Nelson 2015). This breadth of use cases provides evidence regarding the utility of this work for users of cultural heritage resources within Fedora 4 repositories throughout the United States (and even globally).

The performance measurements for API-X will rest upon the agile methodology for software development and upon adoption or development against the API extensions for overall impact. Both DuraSpace and Data Conservancy use an agile methodology for software development that relies upon translation of use cases into requirements that result in specific cards or tasks measured in points. Each time one of these cards is completed, the points are added to the overall "velocity" for fixed period (often two weeks) sprints of software development. The calculation of projected velocity provides a mechanism to organize software development tasks into tractable

and observable clusters while the calculated velocity for each sprint provides a metric and measure for performance.

Being able to meet the requirements and to enable the capability described within the thirty use cases provides a metric for overall impact, especially for the institutions that proposed and developed the use cases. Beyond the original group of institutions working currently on the API-X use case and requirement development, use of the API extensions or development of new API extensions represents another metric for overall impact.

As noted earlier within this proposal, the preliminary work related to use cases has been conducted under the auspices of DuraSpace. All of the API-X outcomes will be published via the DuraSpace wiki, thereby providing a tangible means for documenting outcomes and measuring use and impact. Fundamentally, leveraging the existing resources, processes, practices, documentation, etc. of both DuraSpace and Data Conservancy represents the cornerstone for sustaining the benefits of API-X. The Sustainability and Communications sections of this proposal offer further detail and explanation about this approach.

3. Project Design

The API Extensions Architecture effort began as an informal special interest group with participants from various institutions in the DuraSpace community. The initial activities focused on gathering and analyzing use cases that eventually led to the development of requirements and inspired technical design.

In late 2015, the participants in the API-X effort formalized their roles and identified a set of stakeholders[7] with specific business needs that would be met by API-X with individual extensions. After reviewing the use cases from the broader community, the stakeholders of the API Extension architecture have agreed on an initial list of high-level API-X requirements. Fulfilling these requirements will serve as goals for design and development activities planned for 2016:

1. Infrastructure
    a. The API-X framework shall be deployable on a separate host from Fedora.
    b. It shall be possible to deploy multiple instances of the API-X framework for the sake of scalability of high availability.
    c. The framework shall allow extensions to be hot deployable, configurable, and removable at runtime.
    d. Extension services are exposed as HTTP endpoints.
    e. API-X shall provide a specification that extensions will implement in order to support high level requirements such as service discovery, binding, etc.

[7] https://wiki.duraspace.org/display/FF/People+and+Roles

2. Patterns
    a. API-X shall support a filtering pattern in extensions, where an HTTP request to an existing Fedora URI (or response) is routed through an extension and possibly modified in the process.
    b. API-X shall support exposing new URIs associated with individual repository resources, bound to individual extensions.
    c. API-X shall support exposing URIs associated with the global repository as a whole, bound to individual extensions.
    d. API-X shall support a proxy pattern in extensions, where requests to a URI exposed by an extension are proxied to some other web service.
3. Service Discovery & Binding
    a. It shall be possible to enumerate the list of all extensions deployed and active within the framework.
    b. It shall be possible to enumerate the list of extensions that provide services on a given object in the repository.
    c. For any given object in the repository, it shall be possible to enumerate all URIs exposed by extensions.
    d. API-X shall provide a mechanism for HTTP clients to discover all URIs exposed by extensions on a given object.
    e. It shall be possible to bind extensions to all objects in the repository.
    f. It shall be possible to bind extensions to specific objects in the repository, based on their rdf:type.

Process

At present, the institutions involved in API-X are engaged in informal exploratory activities; members are actively proposing design sketches, proof-of-concepts, and performing background research into existing standards. These activities not only serve to inform concrete design activities for the next steps, but are fulfilling an important communication role by making sure the participants have a concrete shared understanding of the technical challenges implied by the stakeholders requirements and the consequences and tradeoffs of proposed solutions.

As per the typical pattern for community-led development in DuraSpace, formal design and development activities will occur in a series of (typically two-week) development sprints focused around specific milestones that meet stakeholder requirements. Upon completing a set of deliverables in a sprint, the stakeholders and broader community are asked to provide feedback on the sprint results. Given this feedback, the stakeholders and developers agree on a set of milestones for the next sprint. This process proceeds iteratively over the course of the development timeline.

As individual sprints may involve developers from multiple institutions, the timing of individual sprints and the milestones therein are a function of available resources from participating institutions. If awarded, the development resources afforded by this grant will assure that Johns Hopkins' Sheridan Libraries is able to participate regularly and provide leadership and direction to development sprints.

Activities and Timeline

Although the final list of activities for upcoming design or development sprints has not been finalized, the following list of activities is how Johns Hopkins envisions the development process progressing over a six-month timeline. Details of timing for the schedule of completion will be finalized through the DuraSpace led community process:

Proof-of-Concepts (Now ~ March 2016)

- Investigate each of the three proof of concept[8] ideas proposed
- Background research into relevant standards, contemporary practices, and obsolete practices/standards
- Achieve a shared understanding

Fundamentals (March 2016 – April 2016)

- Service binding & Proxy framework components
    - Agree on framework APIs for binding and service discovery
    - Agree on implementation technology
    - Create reference implementations
- Identify extensions to carry forward from proof-of-concepts to fruition as a published extension
    - The intent is to pick one or two 'simple' extensions that fulfill stakeholder use cases, inform development of the framework, and showcase the functionality of the framework once it is publicly available
    - Individual institutions that have business interest in one or more specific extensions will likely provide development resources outside of the scope of this grant, or the API-X effort at large for developing individual extensions

Operationalization (May 2016 - August 2016)

- Demonstrate Hot deployment & configuration
- Devise standards for authoring extensions
- Demonstrate a working 'complex' deployment scenario, e.g. distributed for scalability or high availability

Documentation & Dissemination (August 2016 - December 2016)

- Finalize and publish relevant specifications/APIs
- Documentation for extension authors
- Public release of extension framework to community
- Finish & release any extensions that have been developed
- Present at Open Repositories, Digital Library Federation (DLF) and Coalition for Networked Information (CNI) conferences

As noted within this timeline, the IMLS requested funds would be applied to the Operationalization and Documentation & Dissemination phases.

[8] https://wiki.duraspace.org/display/FF/Proof+of+Concept+ideas

4. Diversity Plan

Not applicable.

5. Project Resources: Personnel, Time, Budget

Sayeed Choudhury, Associate Dean for Research Data Management, will act as project director. Choudhury has been the Principal Investigator for several grants from IMLS, NSF, the Mellon Foundation, the Sloan Foundation, the Library of Congress and Microsoft Research. He acted as PI for the Data Conservancy award and continues to lead its development as product owner. Choudhury's role at Johns Hopkins University specifically focuses on infrastructure development. He has led this work for almost two decades through a combination of grants and base Library funding. The Sheridan Libraries staff includes a budget manager and budget analyst who assists with financial management and grants administration. Both individuals have extensive experience with grants management given the Sheridan Libraries' track record with prior grants. Johns Hopkins University uses an explicit effort reporting system and a wide range of policies and procedures for grants management, all of which have been incorporated seamlessly into the Sheridan Libraries' grants management processes and practices.

Aaron Birkland, Senior Software Engineer, will be the identified software developer supported by the requested grant funds. Birkland has many years of software design and development experience through multiple grant funded projects starting with the National Science Digital Library. He joined the Data Conservancy through the DataNet grant and joined the Sheridan Libraries on a full-time basis last October. Birkland was the principal architect behind the API-X design document that was presented to David Wilcox and Andrew Woods of DuraSpace. Birkland has been playing a leadership role for the API-X bi-weekly telecons. The proposed API-X aligns directly with both Choudhury and Birkland's core duties and responsibilities.

While the Sheridan Libraries in the form of the Data Conservancy initiated this proposed work, the Fedora 4 API-X has become an officially endorsed project of the Fedora community and has enjoyed ongoing participation from organizations such as the Art Institute of Chicago, the Smithsonian Institution, Amherst College, and special interest groups within the DuraSpace community. While the Data Conservancy is prepared to lead the software development, we fully anticipate that other institutions will continue their participation by identifying stakeholders and software developers. This type of engagement is a natural part of the community development process and one of the key benefits of launching this work under the auspices of DuraSpace, which will provide support through its existing collaboration and communication channels such as wikis and conference presentations.

The budget request of $120,500 includes half a year of full-time equivalent salary and benefits for Birkland and one month of salary and benefits for project director Choudhury. The current Johns Hopkins University fringe benefits rate is [redacted]. Additionally, the budget request includes $5,000 of travel support for presentations and community engagement at the Open Repositories, Digital Library Federation, and Coalition for Networked Information conferences. Finally, the budget includes [redacted] of indirect costs applied to all budget categories.

6. Communications Plan

As mentioned in the Project Design section, much of the communications plans rest upon existing processes and mechanisms for any DuraSpace community project. The internal communications amongst the participating institutions and stakeholders has already started. In addition to the focused discussions amongst the participating institutions, API-X has been discussed on the Fedora technical email lists. Choudhury has communicated directly about API-X with two members of the DuraSpace Board of Directors, Robert Cartolano and Tom Cramer. Both Choudhury and Birkland have presented the API-X design and approach to the Library IT Directors from the IvyPlus[9] Universities. Birkland has also presented the API-X project during the Fall 2015 CNI Forum as part of the Fedora Project panel.

Moving forward, DuraSpace and Data Conservancy will continue to use their existing mechanisms of telecons, email lists, websites, wikis, and social media to promote progress regarding API-X on a regular basis. Additionally, the software produced for API-X will be made available through the GitHub Fedora 4 Labs code repository for the entire community.

The budget request includes travel support to the Open Repositories, DLF, and CNI conferences. Given the timing of this grant request, these three events align very well with the timing of API-X development. Open Repositories occurs from June 13-16, 2016 and offers the first public chance to demonstrate results from API-X to a broad audience, including the Fedora committers. The Developer Track represents an ideal opportunity to engage both the technical community and end users about API-X. Given the timing, there would be opportunities to make adjustments or modifications to API-X based on feedback from Open Repositories. The DLF Forum takes place November 7-9, 2016 and the CNI Forum will occur December 12-13, 2016. Given their timing, these events provide an opportunity to demonstrate the final results from the API-X work based on this proposal.

This combination of activities will ensure that there is an iterative process of engagement to influence API-X development beyond the participating institutions. This approach will also ensure that repository managers and software developers from libraries, archives and museums are informed of API-X progress and capabilities to advance their content and data management needs.

[9] The IvyPlus universities include the Ivy League plus Duke, Johns Hopkins University, MIT, Stanford and the University of Chicago. This group has initiated a series of collaborative ventures including BorrowDirect, a program for lending and borrowing books across the entire network and borrowing privileges for visitors from IvyPlus universities.

7. Sustainability

Once the initial production of API-X is complete, we will host it with DuraSpace and leverage their maintenance of products for the community. In particular, we intend for API-X to become an "fcrepo-exts" project. Such projects are considered to be "extension Fedora 4 projects, utilities, and ontologies that are actively maintained." The criteria for such a status are:

1. The module is used by 3 or more installations.
2. The module has two or more formally publicized "owners" (documented in README).
3. The module has an integration-test framework in place.
4. The module has a unit-test framework in place.

The testing requirements (3-4) will be met as a matter of course during development. Through the Data Conservancy, Johns Hopkins University is committed to continued maintenance and development of the framework beyond the scope of the grant. The Sheridan Libraries have committed funding through its base budget to support software developers, a project manager, a business analyst and a system administrator for ongoing development and maintenance of the Data Conservancy. In addition to Johns Hopkins University, the API-X effort currently has five stakeholder institutions with specific business needs that would be satisfied by the API-X framework beyond the timeframe of the grant. Once available for production use, it is highly likely that the API-X framework will quickly meet fcrepo-exts requirements for usage and maintenance and retain that status for the foreseeable future.

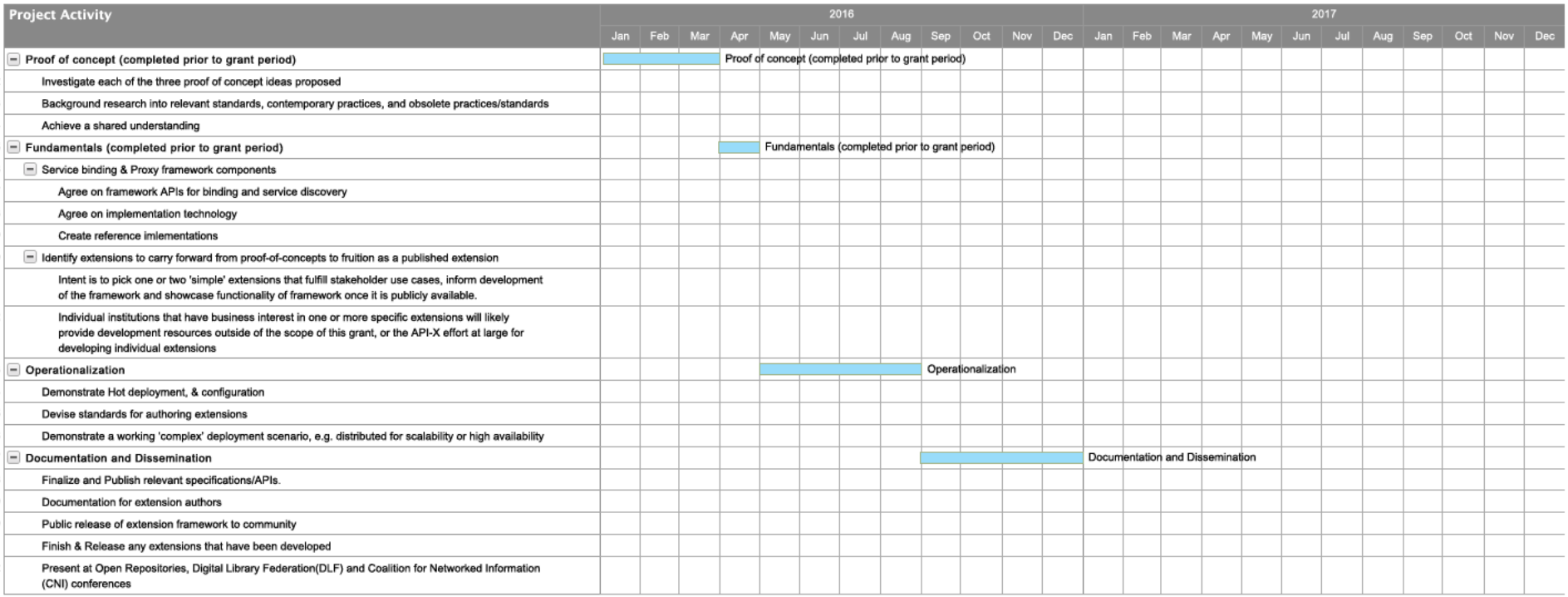
Aside from the framework itself, some individual extensions may share the same fcrepo-ext status, while others may be privately developed by institutions, or even become proprietary. While the development of individual extensions is outside the scope of the grant, a proliferation of API-X extensions maintained by diverse members of the community would help assure its longevity.

A long-term strategic goal of DuraSpace is to transition Fedora from a specific piece of software to a specification of which there are multiple implementations. The decision to instantiate API-X as a separate proxy layer that interacts with Fedora through its HTTP API aligns well with this goal and ultimately reduces risk for adopters as it de-couples their infrastructure from a particular Fedora repository implementation. The initial design process of API-X has already had a synergistic relationship with this effort, as it has raised numerous questions related to the appropriateness of placing functionality into the Fedora software itself, or as an external service.

The sustainability strategy for API-X relies upon the alignment of interests from multiple stakeholders working under the auspices of the community supported processes and framework of DuraSpace. Additionally, the Sheridan Libraries' ongoing support of the Data Conservancy provides additional resources and capacity. If awarded, a grant from IMLS would provide an important and critical source of support to ensure that the initial momentum of API-X matures into a community resource that would be sustained by both Data Conservancy and DuraSpace.

Fedora 4 API-X Architecture

This is the schedule of completion the project team will follow to develop a service proxy layer on top of the Fedora 4 software platform that will facilitate the exposure of repository contents and services as linked data web resources.



Resources Cited Within Proposal Narrative (Listed in Order of Mention)

Data Conservancy -- http://dataconservancy.org

DuraSpace -- http://duraspace.org

DuraSpace API-X wiki space -- https://wiki.duraspace.org/display/FF/Design+-+API+Extension+Architecture

API-X Use Cases -- https://wiki.duraspace.org/display/FF/Use+Cases+-+API+Extension+Architecture

RMap Demonstration Video -- https://www.youtube.com/watch?v=YsZLUdJDFX0

Cool URIs -- http://www.w3.org/TR/cooluris/

Van de Sompel, H., and Michael L. Nelson. 2015. "Reminiscing About 15 Years of Interoperability Efforts." *D-Lib Magazine* 21 (11/12). http://dx.doi.org/10.1045/november2015-vandesompel

API-X People and Roles -- https://wiki.duraspace.org/display/FF/People+and+Roles

API-X Proof of Concept -- https://wiki.duraspace.org/display/FF/Proof+of+Concept+ideas

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets 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 is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets. Your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions
If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project
(Check all that apply):

	Every proposal creating a digital product should complete	Part I
	If your project will create or collect	**Then you should complete**
☐	Digital content	Part II
☑	Software (systems, tools, apps, etc.)	Part III
☐	Dataset	Part IV

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (http://us.creativecommons.org) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

We will license the software under an OSI-approved license so it may be freely reused, distributed, and modified by our target community and beyond. The created software will be publicly available for download, in source and binary forms.

Created software will be licensed under the Apache Software Foundation license (ASL), version 2.0. This license places no restriction on the created software or derivative works, and allows other software projects to reuse the created software without restriction.

Software produced under this grant may be linked by, distributed with, or contributed to by the Fedora community. While many OSI-approved licenses have features similar to the ASL, it is same license used by DuraSpace, Inc, the umbrella agency under which the Fedora Commons operates. We believe the ASL 2.0 license provides maximum legal compatibility with existing and future software artifacts of our

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

Software produced under this grant, including downloadable release artifacts and documentation, will be publically available. The software will be made available according to the best practices of the Fedora community. As of today, this would mean hosting the source code, documentation, and releases on GitHub, a popular, socially-oriented, platform for software development.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

There are no privacy concerns, permissions or rights issues, or cultural sensitivies associated with this proposal.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

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

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

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

B. Digital Workflow and Asset Maintenance/Preservation

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

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, 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 CFR 200.461).

C. Metadata

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

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

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

D. Access and Use

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

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

Part III. Projects Creating Software (systems, tools, apps, etc.)

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) this software will serve.

The Data Conservancy is proposing an extensibility layer between repository infrastructure and the outside web to serve as a sort of HTTP reverse proxy; routing requests public URIs to the services that implement them. This layer will receive HTTP requests from the outside world, perform URL re-writing, potentially modify the bodies or headers of requests and responses, and route requests according to a set of rules or content models present in repository objects. It will serve as a platform for installing and sharing "API extensions" that publicly expose services and features provided by infrastructure that includes a Fedora

A.2 List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

Versions 2 and 3 of Fedora contained a "dissemination architecture" feature which served to bind external web services to Fedora objects. This was originally conceived as a means to associate behavior with repository objects. In a sense, this architecture made Fedora itself serve as a sort of service proxy. This functionality was not initially included in Fedora 4, as the implementation used in Fedora 3 was

B. Technical Information

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

The implementation technologies listed here may change during design and development, if it emerges that other choices may be more suitable.

Java - Java is the most familiar language to a majority of developers in the DuraSpace community, and enjoys a large and mature selection of specifications and third-party libraries that are useful for

B.2 Describe how the intended software will extend or interoperate with other existing software.

Fedora 4 is in the process of defining itself in terms of HTTP APIs layered on top of existing specifications such as the Linked Data Platform (LDP). The API extension architecture plans to leverage these standards as a means of interoperability. In addition, by aligning itself with Fedora's set of HTTP standards, it is envisioned that this infrastructure will be usable with future implementations of Fedora, regardless of

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

The API extension architecture and individual extensions will need to run in a Java Virtual Machine (JVM). Through the use of technologies such as Apache Maven and OSGi, there are automated means of resolving system dependencies with minimal intervention from users or system administrators.

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

DuraSpace maintains a confluence wiki for documentation and development notes related to Fedora 4 development, as well as subprojects and community special interest groups. The working notes from the design and development of the API extensions architecture will be placed onto this wiki, and publicly accessible. Developer and end-user documentation will also be placed onto this wiki, and versioned on a per-release basis as consistent with existing DuraSpace practices.

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

As the development of the API extensions architecture is a collaborative effort between multiple institutions, there is no single sample of code that is representative of the group that will be developing the API extension architecture. That being said, the following are samples of code from developers of the Data Conservancy and DuraSpace community expected to participate in the development process of the API extension architecture:

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? 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 any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

We will license the software under an OSI-approved license so that it may be freely reused, distributed, and modified by our target community and beyond. The created software will be publicly available for download, in source and binary forms.

Created software will be licensed under the Apache Software Foundation License (ASL), version 2.0. This

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

Software produced under this grant, including downloadable release artifacts and documentation, will be publically available. The software will be made available according to the best practices of the Fedora community. As of today, this would mean hosting the source code, documentation, and releases on GitHub, a popular, socially-oriented, platform for software development.

C.3 Identify where you will be publicly depositing source code for the software developed:

Fedora 4 "Labs"

Name of publicly accessible source code repository: GitHub Fedora 4 Labs
URL: https://github.com/fcrepo4-labs

The software developed under this grant is being developed as a modular component of Fedora 4, and is

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

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?

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).

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.

5. What 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).

6. What documentation (e.g., data documentation, codebooks, etc.) 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?

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

8. Identify where you will be publicly depositing dataset(s):

   Name of repository:
   URL:

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

Original Preliminary Proposal

Fedora 4 API Extension (API-X) Architecture

The Sheridan Libraries at Johns Hopkins University request $120,500 to develop a service proxy layer on top of the Fedora 4 software platform that will facilitate the exposure of repository contents as linked data web resources. This work was inspired by a redesign and refactoring of the Data Conservancy software, which was initially prototyped through a grant from the National Science Foundation's (NSF) DataNet program.

Fieldwide Need and Potential Impact

Following the end of DataNet funding, the Data Conservancy decided to apply the lessons learned and redesign and refactor the entire architecture and software platform. The re-imagined Data Conservancy infrastructure builds upon Fedora 4, a digital object repository platform whose development is led by the nonprofit Duraspace[1] organization. While Fedora 4 offers a REST API for basic read/write operations, it does not provide any means to extend this API, or expose services on repository resources. The Data Conservancy software team came to the conclusion that if such a service and extension framework existed, much of the Data Conservancy software platform could be realized as a series of lightweight add-ons to existing Fedora repositories. The Data Conservancy team presented the idea of such a framework to Product Manager and Technical Lead for Fedora and to the community during the Fedora User Group meeting at the 2015 Open Repositories Conference. As a result of these community engagements, this proposed Fedora 4 API extension concept ("API-X") has been adopted as a community project managed under the auspices of DuraSpace.[2]

The first step in this community development process is the collection of use cases. As noted from a link[3] from this wiki space, multiple institutions have submitted twenty-nine use cases as of September 2015. It is worth noting that while the original drivers for API-X relate to the linked data and preservation-oriented approach to data management identified by the Data Conservancy, the range of use cases and interested institutions provides evidence that this proposed work has broader utility and relevance.

Relevance to Agency Priorities

IMLS wisely emphasizes that any national digital platform must leverage and extend existing technology. There is an existing user base of Fedora software for institutional repositories that will grow given the important enhancements and robustness offered by Fedora 4. By providing an architecture to deploy repository services as lightweight extensions, institutions that use Fedora 4 for their institutional repository needs would be

[1] http://duraspace.org

[2] https://wiki.duraspace.org/display/FF/Design+-+API+Extension+Architecture

[3] https://wiki.duraspace.org/display/FF/Use+Cases+-+API+Extension+Architecture

automatically positioned to extend their platforms for more robust data management. As federal funding agencies response to the White OSTP memoranda regarding public access to publications and data, it is becoming clear that simply depositing and subsequently downloading data will not be sufficient. The proposed work supports a vision of data management where data are packaged with information graphs that capture and preserve connections to publications and software.[4]

Project Director and Partners

Sayeed Choudhury, Associate Dean for Research Data Management, will act as project director. Choudhury has been the Principal Investigator for several grants from IMLS, NSF, the Mellon Foundation, the Sloan Foundation, the Library of Congress and Microsoft Research. He acted as PI for the Data Conservancy award and continues to lead its development as product owner. While the Data Conservancy initiated this proposed work, the Fedora 4 API-X has become an officially endorsed project of the Fedora community and has enjoyed ongoing participation from organizations such as the Art Institute of Chicago, the Smithsonian institution, Amherst College, and special interest groups within the Duraspace community. While the Data Conservancy is prepared to lead the software development, we fully anticipate that other institutions will continue their participation by identifying stakeholders and software developers.

Proposed Work Plan and Outcomes

At present, the API-X effort is collecting and analyzing use cases for potential extension modules from the community at large. Near the end of October, we expect to formally designate individuals as stakeholders or developers. The stakeholders will agree on an initial set of requirements for the framework, and an initial set of extensions. Development will follow an agile methodology. The stakeholders will define a set of tasks that address the requirements, and the developers will schedule intense two-week 'sprints' to produce a tangible product milestones that will be evaluated by the stakeholders. This process will repeat iteratively until all requirements are met, and the framework is released.

Budget

The budget request of $120,500 includes half a year of full-time equivalent salary and benefits for a software developer and one month of salary and benefits for project director Choudhury. The current Johns Hopkins University fringe benefits rate is [redacted]%. Additionally, the budget request includes $5,000 of travel support for community engagement at the DLF, CNI and Open Repositories conferences. Finally, the budget includes [redacted]% indirect cost applied to all categories of budget categories.

[4] For example: https://www.youtube.com/watch?v=YsZLUdJDFX0