Fedora Commons

Proposal to the Gordon and Betty Moore Foundation

Submitted by Sandy Payette

Table of Contents

1 Fedora Commons: Vision and Motivation

The purpose of this proposal is to request four years of funding for the startup of Fedora Commons, a non-profit organization that will be the home of the existing Fedora Project[1]. The mission of Fedora Commons is to provide the technical expertise and community framework to develop sustainable open-source software to support innovative collaboration and knowledge sharing among scientists, scholars, and educators, while ensuring the integrity and longevity of the results of their work. As described throughout this proposal, the carefully coordinated community of technical and domain experts that we propose in Fedora Commons is integral to the goal of creating such a software framework.

The World Wide Web has unleashed unprecedented sharing and dissemination of information. The recent phenomenon of "Web 2.0" [22] takes the next step by turning the Web into a distributed platform for collaboration and participation. Web applications have evolved to enable the formation of large interactive communities that share both ideas and objects. Anyone can easily participate in these communities, for example by collaboratively authoring articles in Wikipedia, sharing photos on Flickr, writing reviews of products on Amazon, and posting videos on YouTube. The essence of this phenomenon was articulated by Tim O'Reilly, who observed that "the central principle behind the success of the giants born in the Web 1.0 era who have survived to lead the Web 2.0 era appears to be this, that they have embraced the power of the web to harness collective intelligence." [22]

Scientists and scholars are now beginning to realize that such social networking and information sharing technologies have the potential to make revolutionary changes in the process, pace, and quality of intellectual discourse and scientific results. Consider the potential of applications like the new PLoS ONE[2] system, one of our partnerships, that give scientists open access to scientific literature and provides mechanisms that allow experts to discuss, review, and refine these results online.

Similarly, educators and museum curators are realizing that the social web that fills much of the leisure time of students is a paradigm for improving the process and quality of education. Collaborative media sharing applications such as YouTube and Flickr provide a model for new educational applications that can, for example, enable the sharing of the latest images from Hubble and facilitate the use of web-based tools for experimentation based on these images. Our partners in the National Science Digital Library[3] are developing such tools in collaboration with the Exploratorium museum.

But, what happens if these new collaborative environments result in information becoming locked up in a new wave of clever, yet idiosyncratic, systems that are not built to facilitate sharing across their individual boundaries, and are not attentive to the long-term sustainability of social and intellectual knowledge accumulated within them? This is where Fedora Commons can play a critical role. We argue that this new wave of applications should be constructed with *interoperability* in mind. For example, a teacher using a curriculum authoring application should be able to combine Hubble tools

[1] http://www.fedora.info
[2] http://www.plosone.org/home.action
[3] http://nsdl.org

from the Exploratorium, raw datasets from the National Virtual Observatory[4], and visualization tools created by the Cornell Theory Center[5]. Furthermore, this teacher and her students should be able to, as part of the use of this curriculum, engage in collaborative discussions with scientists. In addition, these applications should be constructed with *preservation* in mind. The information created within these applications – papers, data, annotations and commentary – is part of the scholarly and cultural record and should be well-managed so that they persist over time.

Fedora Commons will embrace the ***dual focus*** of enabling the creation of innovative, collaborative information spaces *and* attending to the longevity and integrity of information that results from collaboration. In doing so, it will work to ensure that change is both revolutionary and sustainable. Fedora Commons builds on the foundation of the highly successful and internationally recognized Fedora Project[6]. The unique contribution of Fedora was cited by Reagan Moore, Director of Data and Knowledge Systems at the San Diego Supercomputer Center and an internationally recognized expert in digital preservation and e-science. At an Oxford Institute conference on digital preservation and super-computing, Dr. Moore identified Fedora as an example of a system that brings together the strands of ontology driven semantic interoperability across heterogeneous collections, dynamic linking, and digital preservation in institutional repositories [31].

The existing Fedora software has been adopted by hundreds of institutions for an array of innovative applications in education, open-access publishing, scholarly communication, and e-science. Among its notable installations are the Topaz/PLoS ONE open access journal system, the National Science Digital Library (NSDL), the Max Planck Society's e-scholarship system[7], the Chicago Historical Society's multimedia encyclopedia[8], the University of Virginia's digital collections[9], the Australian national institutional repository initiative (ARROW)[10], Oxford University's digital archive[11], the Perseus humanities computing project[12], and numerous applications in national libraries, companies, universities, and cultural institutions. This deployment of Fedora to this array of institutions has created a large "Fedora Community", evidence by six Fedora User Conferences over the past two years in the U.S., Europe, and Australia.

Our plan over the next four years is to build on the trajectory established by the Fedora Project and build a self-sustaining Fedora Commons organization. We plan to do this with a three-pronged effort that is described in detail in the body of this proposal:

[4] http://www.us-vo.org/
[5] http://www.tc.cornell.edu/
[6] http://fedora.info
[7] http://www.escidoc-project.de/homepage.html
[8] http://www.encyclopedia.chicagohistory.org/
[9] http://www.lib.virginia.edu/small/online_res/digital.html
[10] http://arrow.edu.au/
[11] http://www.hcu.ox.ac.uk/jtap/
[12] http://www.perseus.tufts.edu/

1. *Technology Development* – We have defined a number of areas where the current software base will be enhanced and functionality expanded including improved facilities for integration of collaborative technologies, greater leveraging of semantic web technologies, integrations of enterprise capabilities such as workflow, and attention to the vital preservation worthiness of the architecture.

2. *Community Development* – As we describe later, one key to an effective open source project is an active, large, and enthusiastic community of users and technologists. These have been characterized elsewhere as "passionate users" [25]. A significant part of our effort therefore will be dedicated toward community building so that at the end of funding there will be a sufficient community network effect to carry Fedora Commons forward.

3. *Leadership* – As Steven Weber notes [35] an open source project can descend into chaos without coordination from the top. The leadership team of Fedora Commons will combine the vision and experience of the existing Fedora Project, a perspective on both industry and web trends, and expertise in outreach/communications, all essential to the success of Fedora Commons.

In closing, we view Fedora Commons as an excellent complement to the majority of the Moore Foundation's funding, which focuses on Environmental Conservation and Science and funds researchers collecting data in many remote areas of the world. The Fedora Commons open-source software platform will enable these communities of scientists, researchers, and educators to more easily integrate their digital resources, capture relationships among these entities, reuse parts of them in new contexts, and allow easy integration of digital information across community and institutional boundaries. Like other open source projects the impact can extend to both the developed and the developing world [15].

2 The Power of Open Source Software

Fedora Commons will be an *open community* to support the development and deployment of *open source* software, which facilitates *open collaboration* and open *access* to scholarly, scientific, cultural, and educational materials. This *open world*[13] is a phenomenon of the networked information environment that has been embraced by scholars, entrepreneurs, and politicians due to its economic, technical, political, and cultural benefits. Chris Kelty in his forthcoming book *Two Bits: The Cultural Significance of Free Software*[16] argues that FOSS (Free and Open Source Software) is an example of a broad shift in the structure of power, technology, markets, and corporations. In the remainder of this section we summarize the importance of openness and its relevance to the organization and mission of Fedora Commons.

[13] This term, which covers notions of open collaboration, open community, open access, and open source was coined by Shay David, a Ph.D. candidate at Cornell [10].

Yochai Benkler explores the economics of the "networked information economy" in his recent book *The Wealth of Networks* [3]. He argues that this new economy is based on the mass distribution and rapidly dropping cost and proliferation of information producing and sharing technologies, enabling a model of *peer-production.* Benkler notes the efficiency and sustainability of the open world due to its basis in "commons-based peer production", a phenomenon we see manifested in software projects such as the proposed Fedora Commons effort. In this model the incremental effort of many individuals results in products that are recognized as superior to their commercial counterparts, as demonstrated by the high quality of open source software and its widespread deployment in "mission critical" environments. The efficiency of this open source strategy has been recognized by a number of technology companies such as IBM [4], Sun [27], and Red Hat [28]. Our own work has benefited from collaborations in both the non-profit and commercial sectors. A specific goal of Fedora Commons is to develop many new partnerships. For example, we are currently investigating a collaboration with Sun Microsystems related to archival storage solutions.

Steven Weber in his book *The Success of Open Source* [35] describes that the factor that prevents open-source code from becoming a mess is a sophisticated governance structure and editing function so that the code, over time, works together and is maintainable. Thus, a key part of this proposal is funding for the organization and management of the Fedora Commons community, which as described later will need to balance the "controlled anarchy" of the commons production model and the management required to produce usable products.

Another economic point that Benkler [3] makes is the inconsistency and inefficiency of market-based *proprietary* information strategies when applied in *non-proprietary* sectors, such as the education and scholarship domains that are particular focus areas of the Fedora Commons. He describes that proprietary software imposes an unfair burden upon these sectors that base their financial strategies largely on non-proprietary government and foundation funding.

A number of observers of the open world phenomenon point to the beneficial cultural and political benefits of open source and open access. These include Gabriella Coleman [8] who describes it as an extension of free speech and expression, and Lawrence Lessig [18, 19] who sees it as critical for freedom of exchange of ideas that is critical to democracy. The UN recognizes these factors, as stated in the UNESCO-sponsored publication *Breaking Barriers. Using free and open source software for development* [15]: "the benefits offered by FOSS have been particularly useful for poor developing countries around the world … [making] information and communications technology (ICT) more affordable to them … [thus] bridging the digital divide." Similarly, Charles Vest, President Emeritus of MIT, writes about the importance of open source and open content for a *meta-university,* a global learning environment that will allow students in developing countries to access and use learning and research materials unavailable in their locale [33].

From the perspective of higher education, Brad Wheeler CIO of Indiana University [36] states that "The Community Source Model for developing and sustaining software is a remarkable fit to the culture and values of higher education." Szulik from Red Hat [28] observes that the open source model has its

origins in the traditions of learning institutions around the world. It is built around the premise that technology is transparent, allowing the best ideas to win in an open, collaborative environment, as well as the idea that open source development should contribute to a body of knowledge that others can learn from and build on.

Fedora Commons fits in this "open world" context that according a wide range of educators, scholars, and policy maker is increasingly critical to our future. With the support of the Moore Foundation we will be able to build on our success thus far and make Fedora Commons a model of all that is best about openness.

3 Leveraging the Success of the Existing Fedora Project

Fedora was founded as the Flexible Extensible Digital Object Repository Architecture and is the result of DARPA and NSF-funded research led by Sandy Payette and Carl Lagoze at Cornell University in the late 1990s [23]. The primary motivation of this research was to devise new information architectures to facilitate interoperable access and management of the increasingly complex and heterogeneous digital collections. These collections were rapidly emerging as the result of (1) massive scanning efforts to put the world's knowledge online in vast "digital libraries" and (2) the introduction of new technologies into the scientific and scholarly process resulting in "born digital" materials which were quickly becoming the foundation of the new scholarly record. With each new project devising its own idiosyncratic solution for dealing with digital material, the possible future was an unmanageable set of stove pipe systems that would inhibit interconnection of information and endanger the long-term welfare of digital resources. Thus, a primary goal of the initial architecture was to design a uniform "digital object" model that could represent the full variety of digital content and a generalized repository model for consistent access and management to content.

This notion caught the attention of researchers at the University of Virginia Library [26] who were evaluating solutions for managing an increasingly diverse digital collection and digital humanities corpus[14]. After a period of successful experimentation with prototypes, the University of Virginia and Cornell University jointly applied for funding from the Andrew W. Mellon Foundation to build a production-quality, open-source system based on the original Fedora architectural principles. Mellon has funded the Fedora Project since 2001 and at this point version 2.2 of the open source software is available. The enduring first principle of a common object model over diverse content remains, but has been enhanced by rich new features including integration with web services, semantic web relationships, fine-grained access control, and version management. These features are implemented as a service-oriented architecture that is highly scalable and robust, and that integrates with other applications and services.

The three most important themes of the Mellon-funded era of the Fedora Project are described below.

[14] http://www.iath.virginia.edu/

3.1 Flexible Building Blocks to Create the New Digital Landscape

The flexible and extensible *digital object* model at the core of Fedora can be characterized as providing "Lego-like" building blocks to support uniform management and access to heterogeneous content including books, images, articles, datasets, multi-media, and more. The model also supports the aggregation of content from distributed network sources into new complex digital objects, and the construction of information networks by inter-relating multiple digital objects.

The flexibility of the object model makes it possible to compose complex content in a variety of configurations. One example is depicted in Figure 1, with three Fedora digital objects from the special collections library at the University of Virginia. The figure shows three independent digital objects, one with the encoded text of a historical document, another with the page scan of the document, and another with a relevant art image. Each object has its own metadata. Although each object is an independent entity with its own persistent identifier, the objects can be related to form a compound digital object.

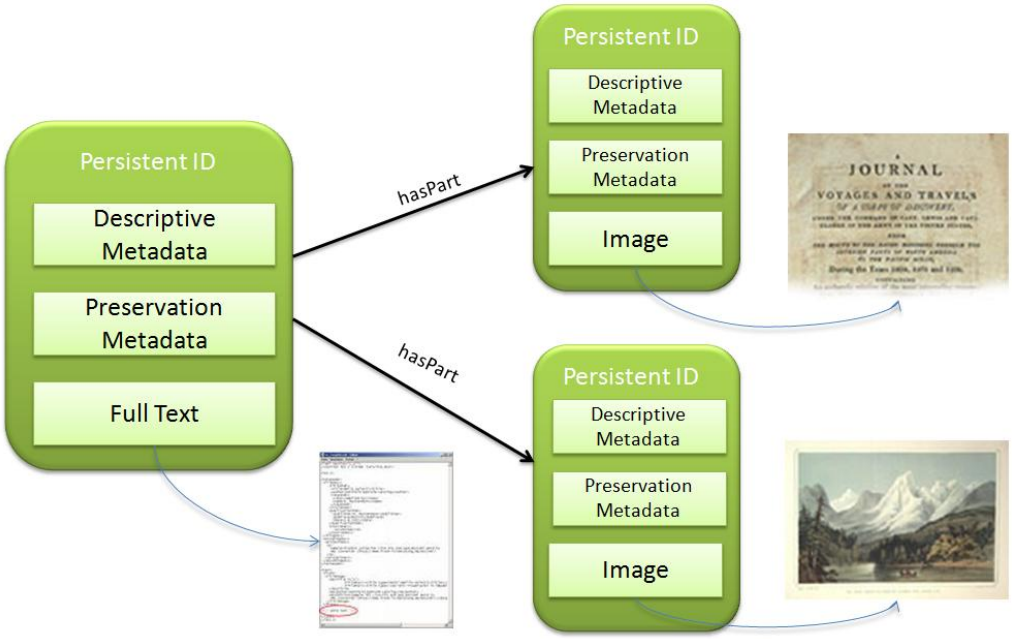

Figure 1 - Historical journal text and related art image

In addition to providing the building blocks for content composition, Fedora also includes a generic repository *service* that defines and implements the core set of functions and interfaces necessary to manage and disseminate complex and heterogeneous digital objects in a generic and interoperable manner. The repository service allows the binding of services to objects to create dynamic views of digital content. An important aspect of Fedora is that its digital objects are self-describing, which means that they encapsulate all essential characteristics of the object. This allows a complete rebuild of the repository from the information in the core objects, which has significant benefits for preservation and archiving of digital objects. This is a subject described in section 6.4.

3.2 Catching the Next Big Wave - Semantic Technologies

Semantic technologies are viewed by many to be the next major wave of technology in the information economy. These technologies have a central role in systems that support knowledge work. This impacts not only new systems but also existing systems that must be poised to adapt and evolve to integrate semantic technologies as they mature.

In 2005, the Fedora Project successfully integrated a first stage of semantic web technologies into its repository service by introducing support for RDF descriptions [17] of digital objects, RDF relationships and graph-based indexing of Fedora repositories [15]. This provided a more powerful and generalizable mechanism for expressing the relationships among digital objects. With this implementation, the objects in the repository could now be treated as nodes on a graph and that graph could be queried. The impact of this adaptation has been significant since it positions Fedora as a unique open source system that simultaneously combines rich, semantic "information networks" of digital content with support for persistence and management of digital content in a robust repository system that includes features such as authentication, policy enforcement, versioning, replication, integrity checking, dynamic views of digital objects, and more. Figure 2 shows a simple example of the use of semantic relationships, demonstrating both the connection of a digital object to its Dublin Core metadata properties and to its component parts, an abstract and the full-text of the article.

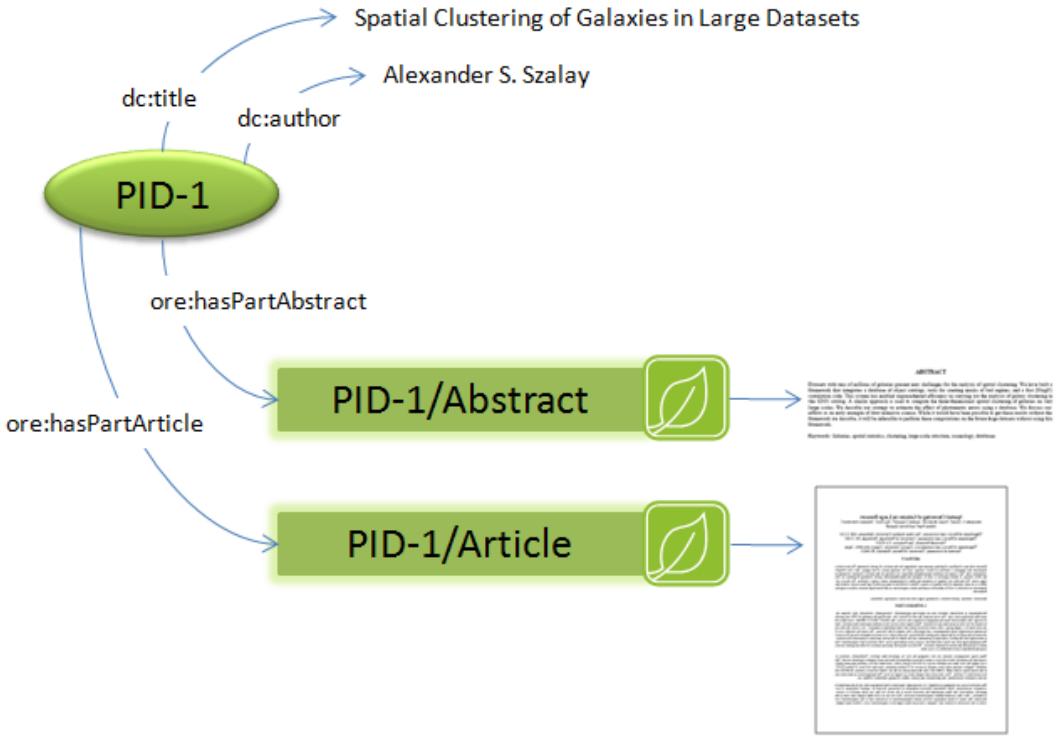

Figure 2 - Article and abstract as graph

[15] http://www.fedora.info/download/2.2/userdocs/server/resourceIndex/index.html

Figure 3 then expands this simple example to connect the digital object to other digital objects, one of which is a review of the article and the other a data set associated with the article. Because these are all nodes in a graph, these component digital objects can be readily reused in other contexts. In addition the richness of the relationships in this graph can be queried and digital object content can be accessed and modified via well-defined Fedora APIs.

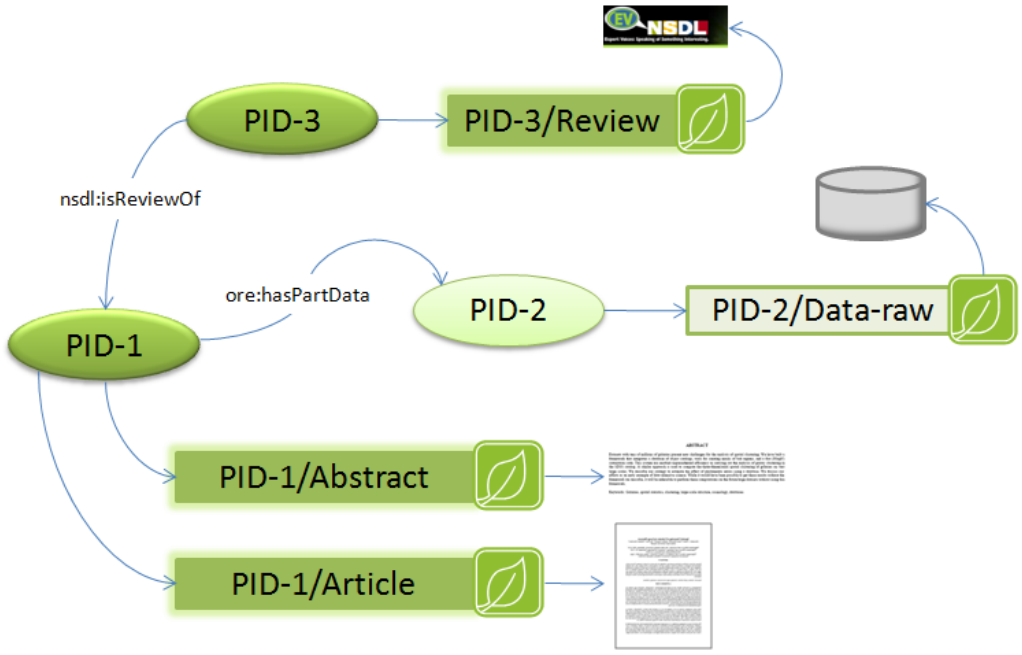

Figure 3 – Adding data and reviews to the graph

3.3 Existing Fedora: An Agile System with Enduring First Principles

As shown in Figure 4, the existing Fedora system is logically divided into four major functional areas that reflect its first principles: repository services, preservation services, semantic services, and enterprise services. We argue that these are critical services that should be offered by a platform whose purpose is to enable collaborative applications while attending to the challenges of information management and preservation.

The Fedora open source system is designed to permit alternate configurations. Also it is designed with the assumption that technology will change and the system will have to adapt and evolve. Overall, Fedora is an agile system that can be used in many ways, and can adapt to new requirements as they emerge.

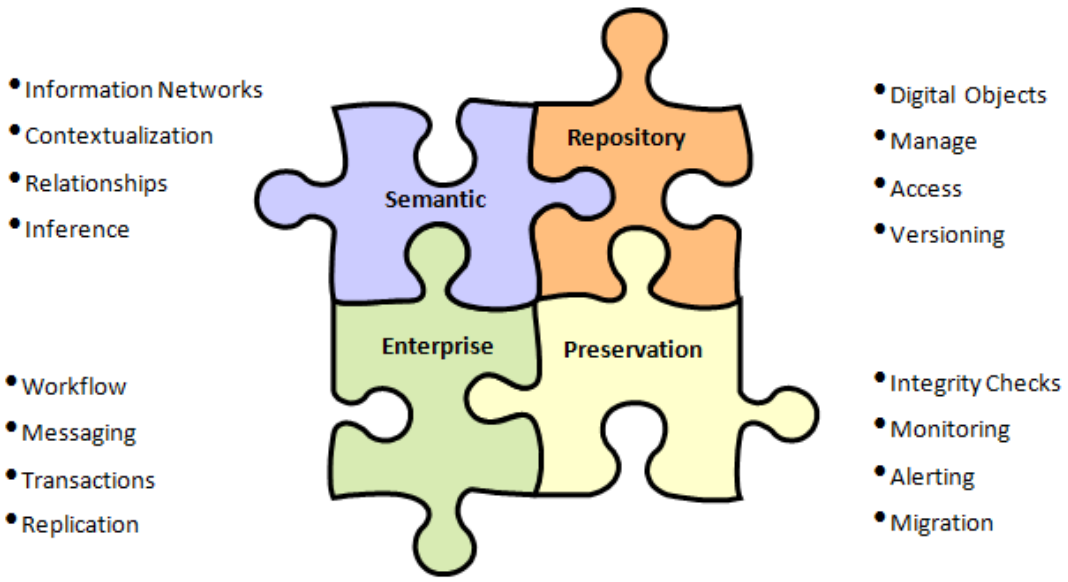

Figure 4 - Fedora: integration of key services

4 Investing in the Future: Fedora Commons Community and Technology

Fedora Commons will leverage the success of the existing Fedora Project and create the organizational and technical frameworks necessary to support revolutionary change in how scientists, scholars, and educators produce and share their intellectual outputs, and ensure the integrity and longevity of information. There are two main thrusts to its mission: *community* and *technology*.

Section 5 describes the first thrust which is to create and nurture a committed and thriving community that collaborates in evolving, maintaining, and using the software. A core Fedora community already exists on an international scale as evidenced by the number Fedora user group meetings worldwide, and in Fedora being featured in prominent conferences such as World Wide Web 2007, Open Repositories 2007, and the European Semantic Web Conference 2007. The Fedora Commons will engage this existing community and begin a substantial outreach effort to build new community.

Section 6 describes the second thrust which is to create a technical framework that enables innovation, development, management and deployment of high-quality open-source software. Fedora Commons will provide leadership to coordinate distributed community-based software development and to integrate new innovations into the open-source platform in a manner that is consistent with the overall Fedora Commons vision and with new directions in the Web and enterprise systems.

5 Community Strategy: Development of High-Impact Community Partnerships

Fedora Commons will devote significant effort to the development of high-impact partnerships and community outreach in four key community sectors: (1) scholarly communication, including open

access publishing, (2) e-science, (3) education, and (4) libraries, museums, cultural institutions. As depicted in Figure 5, while each community is distinct in its own right, these communities also interact with each other in many ways. Interesting Fedora-based systems are already being built that integrate digital information across these different community sectors.

Figure 5 - Four target community sectors for Fedora Commons

As part of the Fedora Commons start-up phase, the Fedora Commons leadership will focus on strategic partnerships with an initial group of highly committed organizations and projects. Each organization holds a highly-visible leadership position in its respective community sector, and each is already building innovative systems based on the Fedora open source software. These institutions recognize that Fedora Commons will provide the essential organizational context to foster innovation, to integrate new technologies into the Fedora Commons software framework, and provide a focal point for collaboration among information-building communities that are in pursuit of a common vision. The exact definition of "partnership" will be developed jointly with these organizations (consistent with the Fedora Commons articles of incorporation and bylaws). Initially, the partnerships will manifest as strong alliances in which organizations work in close strategic alliance with Fedora Commons leadership.

An anticipated outcome of these partnerships is that "flagship" systems will emerge that will serve as reference models to the broader communities. We anticipate that these systems can either be directly redeployed as open source, or be easily re-created by others using the Fedora Commons open source technologies. Also, partners will play a significant role in defining new system requirements for Fedora Commons to guide the evolution of the open source software. Finally, partners will help the Fedora Commons leadership formulate its outreach strategy and provide key contacts into the target community sectors.

It is recognized that the work of the partner organizations often crosses the boundaries of more than one target community sector. The Fedora Commons leadership will facilitate interactions across the sectors and will ensure flexibility of information flow and collaborative possibilities. The table below lists initial strategic partners and their associated target sectors. The sections that follow describe the significance of each partnership.

Initial Strategic Partners	Scholarly Communication and Open Access Publishing	E-Science	Education	Libraries, Museums, Cultural
PLoS ONE and Topaz	☑	☑		
eSciDoc (Max Planck Society and Fiz Karlsruhe)	☑	☑		
Johns Hopkins University and National Virtual Observatory	☑	☑		
Australian ARROW and DART	☑	☑		
National Science Digital Library			☑	☑
OhioLINK Digital Commons	☑		☑	
Tufts University and Perseus	☑			☑
University of Virginia *	☑			☑
Cornell University *	☑	☑		

* Fedora founding institutions (will provide both a research and development focus)

5.1 Scholarly Communication and Open Access Publishing Sector

In the open-access publishing sector, we are forming strategic partnerships with Topaz and the Public Library of Science (particularly the PLoS ONE initiative). PLoS ONE brings a new dimension to open access publishing by deploying novel Web 2.0 applications that promote user participation. Specifically, PLoS ONE can empower scientific communities to engage in dialog on papers and articles in a way that has not been possible before through the use of threaded discussions, annotations, post-publication ratings/reviews, wikis, tags, user customizations and computing usage statistics. Taking advantage of

the full potential of the Web, PLoS ONE is positioned to fundamentally change how scholarly journals will be used, and it can enable a powerful and interactive process of network-based scientific and scholarly collaboration.

Fedora Commons will work with PLoS ONE to define new systems requirements for evolving the Fedora Commons open source software to best enable these new modes of dissemination and sharing of scientific information. Also, Fedora Commons will consult with PLoS ONE in devising a community outreach plan to propagate the open access philosophy in general and the Fedora Commons technologies that support it in particular.

Topaz is developing the underlying application framework for PLoS ONE. The Topaz framework is built upon Fedora and Mulgara[16] (an open source triple store technology). Topaz is developing an end-to-end electronic publishing system that will become a powerful driver of the open access movement. The Topaz application framework is introducing a set of innovative tools and resources to enable access, use and analysis of scientific literature and data. While, PLoS ONE will be the first user of the Topaz application framework, the Topaz outputs will be made available in open source to facilitate the shift of the broader scientific and medical communities from subscription based journals to open access publications.

Through its partnerships with Topaz and PLoS ONE, the Fedora Commons technical leadership will strategize on how to move the Fedora Commons software in a direction that best promotes innovation in the scholarly communication and open access publishing sector. We propose that Fedora Commons become the future home for open source components developed by Topaz. Fedora Commons will work closely with Topaz to devise a system integration plan that enables the Topaz open source components to become part of the Fedora Commons open source software offering. A significant aspect of this system integration will be to determine the best ways to improve Fedora's integration with Web 2.0 and semantic technologies. The Topaz team is already investigating a compelling architectural pattern that could offer the option of a repository that is more natively graph-oriented. This technical integration is discussed further in Section 6.

Another strategic partner in this sector is eSciDoc[17], a joint project of the Max-Planck Society[18] and FIZ Karlsruhe[19] in Germany. eSciDoc is an ambitious project to support collaborative work among scholars distributed across Max Planck laboratories throughout Germany, provide new models of information access for these scholars, ensure permanent access to the research results and research materials of the Max-Planck Society, and integrate these materials into an emerging, global, electronic knowledge space. In January 2006 the management of eSciDoc made the descision to use Fedora as the technical framework of the project[20]. This decision was made after careful analysis of a number of commecial and

[16] http://www.mulgara.org/
[17] http://www.escidoc-project.de/homepage.html
[18] http://www.mpg.de/english/
[19] http://www.fiz-k.com/
[20] http://www.escidoc-project.de/news.html

open source alternatives. Since that decision, we have worked closely with the eSciDoc team to understand how to best evolve Fedora and to integrate eSciDoc contributions into the Fedora system. In addition, the eSciDoc team has been an active participant to the Fedora Community, speaking at workshops and participating in architectural planning.

The eSciDoc partnership is particularly valuable since the project is concerned with issues that are critial to the evolution of Fedora to an enterprise-level application. These include versioning, scalability, and system availability. Our work thus far with eSciDoc has been of tremendous benefit in understanding the issues in these areas and we look forward to the fruits of this partnership under the Fedora Commons umbrella. A letter of support from eSciDoc management is attached to this proposal.

Another key partnership in the scholarly communication sector is ARROW (Australian Research Repositories Online to the World)[21]. ARROW has been funded by the Australian government since 1993 and represents a commitment in Australia to make scholarly results available online at a large scale. At this point sixteen of the forty universities in Australia have adopted ARROW making it one of the premier institutional repository projects in the world. The software platform for ARROW is Fedora. This choice was made based on the flexibility of Fedora, its rich feature set, and the international community of users who participate in the development. ARROW has agreed to be an enthusiastic partner in the Fedora Commons. Based on their active participation in the Fedora Community over the past several years this promises to be a valuable partnership. Andrew Treloar, Chief Architect both the ARROW and DART projects is also a current member of the Fedora Project advisory board. *A joint letter of support from Dr. Treloar and the University Librarian at Monash University is attached to this proposal.*

5.2 E-Science Sector

In the e-Science sector we are forming a partnership with two significant players in this area: Johns Hopkins University (JHU) and the Dataset Acquisition Accessibility & Annotation e-research Technologies (DART) Project[22].

Our partnership with JHU is in the context of its close affiliation with the NSF-funded National Virtual Observatory (NVO)[23]. NVO is frequently cited as a model for what is referred to in the U.S. as "cyberinfrastructure" [21] and e-Science in the UK and Europe. Stated very briefly, both terms refer to data and computationally-intensive science that exploit open standards, networks, and distributed high-speed computing. The NVO has been funded since 2001 by the NSF and over the years its work has exemplified, within the astronomy community, the notions of the "global information commons" and "flexible technical architecture using standard, open protocols and interfaces" that the NSF describes in its cyberinfrastructure vision report [21, 37].

A key challenge in data-intensive science and scholarship is *data curation*. Again, quoting the NSF report [21]: "The anticipated growth in both the production and repurposing of digital data raises complex

[21] http://arrow.edu.au/
[22] http://dart.edu.au/
[23] http://www.us-vo.org/. The Principal Investigator and Project Manager of NVO are located at JHU.

issues not only of scale and heterogeneity, but also of stewardship, curation and long-term access." This is especially critical in a new vision of publication that has become increasingly widespread whereby scholarly documents directly incorporate datasets and simulations based on those datasets [5, 6, 32, 34, 37]. This vision depends on the longevity of the datasets incorporated into publications.

The leadership of the Digital Knowledge Center (DKC)[24], which is part of the Sheridan Libraries at JHU, has been working closely with NVO to model distributed data curation technologies [12]. This work is supported by JISC, IMLS, Microsoft, SPARC, and TeraGrid. Based on a Mellon-funded study comparing features of performance of various repository implementations [7], the DKC selected Fedora as the foundation technology for its prototype data curation system. Their decision was made based on a number of characteristics of Fedora:

- As established by JHU testing [7], Fedora's performance and robustness is well-matched to the demanding tasks required to support the curation of large data sets over long periods.
- Fedora is open source software which is essential because the JHU team must have the freedom to modify and redistribute the Fedora software as part of its data curation system.
- Fedora enables the association of behaviors, or "disseminations," with digital objects which enable dynamic transformation and adaptation of digital content so that it can be used in multiple application contexts.

We expect that this valuable partnership link the Fedora Commons effort to the large e-science community. It will also ensure that Fedora Commons technology evolve consistent with this community's needs. There are plans to significantly expand the JHU partnership in the context of a major new proposal being formulated for the upcoming NSF cyberinfrastructure call. *A letter of support from Sayeed Choudhury, Director of JHU's Digital Knowledge Center, is attached to this proposal.*

The DART project [30] is a major collaborative project funded by the Australian government to investigate infrastructure for e-Science. This project builds on the work of an earlier project ARROW [29], which developed institutional repository infrastructure based on Fedora. Based on the experiences with their deployment of Fedora in that earlier project, and on the unique feature set of Fedora, DART has chosen to use Fedora as a central part of the DART technology. As noted by Andrew Treloar, project architect of DART, "The digital objects that DART stores need to be managed, preserved, persistently identified, aggregated and disseminated in flexible ways" and "Fedora has been found by a range of projects to be the best match for these requirements" [30].

5.3 Education Sector

In the education sector, Fedora Commons has developed a strategic partnership with the National Science Digital Library (NSDL) which is focused on enabling radical changes in math and science education by creating a new "Web 2.0" world upon Fedora-based digital repositories and services.

[24] http://dkc.jhu.edu/

The NSDL is a major initiative funded by the National Science Foundation (NSF) [38] with the goal of improving education through innovative uses of new technologies. In 2000, NSDL created of an online digital library to direct users to exemplary resources for science, technology, engineering, and mathematics (STEM) education and research. This digital library has evolved into a collaborative, semantic digital library that provides services and tools that enhance the use of STEM resources in a variety of contexts. NSDL is designed primarily for K-16 educators, but anyone can access and search the library[25].

With over 2.5 million resources managed and deployed through its Fedora-based repository - and with a social network of educational organizations, universities, publishers, learned societies, and collaborators – the NSDL serves as a new kind of information commons where bridges are being built between:

- private sector and public interests by providing access to resources such as publishers' journal articles, teacher-created lesson plans and real-time data sets from scientists
- scientific and educational communities by applying advanced technologies to stimulate new ways for educators and learners to access and use scientific information
- teachers and students at all levels, in all locations by supplying content and tools in open-access, non-proprietary formats in an easily accessible online environment.

The NSDL will play a significant role as a strategic partnership of Fedora Commons. The NSDL can be viewed as a "hub" that connects Fedora Commons into the broader education community. Through the NSF-funded Pathways program[26], the NSDL has ten partners who, in turn, collaborate directly with 85 projects, organizations and societies in the disciplines of Chemistry, Biology, Computational Science, Engineering, Materials Science, Physics and Astronomy, Middle School Education, Community and Technical College Education, and Multimedia Education. Further NSDL is recognized in NSF's cyberinfrastructure plans for its potential to spread data-intensive scholarship to the education sector [21]. We can leverage the important connections the NSDL has already made as part of the Fedora Commons outreach effort. *A letter of support from Kaye Howe, the Executive Director of NSDL Core Integration, is attached to this proposal.*

5.4 Library, Museum, and Cultural Sector

With roots in the early digital library work of the 1990s, the Fedora Project formed many long-standing relationships that will be the basis of important strategic partnerships for Fedora Commons.

Fedora Commons will benefit greatly from its strategic partnership with University of Virginia (UVa), which is one of the founding institutions of the Fedora Project. As previously mentioned, UVa co-developed the Fedora open-source software with Cornell and has provided the development team with a deep understanding of the challenges of managing and disseminations complex digital collections.

[25] http://nsdl.org
[26] http://nsdl.org/about/?pager=pathways

The UVa Library has a tradition of early innovation and adoption of technologies to provide rich digital collections to scholars, and innovative tools for using those collections. UVa has played a unique role in Fedora as being both one of its developers and one of its major users – a very nice example of "drinking one's own champagne."

Interestingly, the legacy of Thomas Jefferson has served as an inspiration in UVa's new effort to create an integration academic knowledge environment.

> "***The field of knowledge is the common property of all mankind***."
> Thomas Jefferson, 1807, Correspondence with Henry Dearborn, U.S. Secretary of War

In 1819, when Thomas Jefferson founded the University of Virginia, he envisioned an "Academical Village," where students, scholars and other community members could pursue knowledge together. At the beginning of the 21st century, the University finds itself poised to become a new embodiment of that integrated academic environment, characterized by technology-enabled modes of information creation and consumption. More and more universities and libraries are supporting faculty and students as they create elaborate research and teaching materials that involve all manner of digital resources, and use these materials to collaborate with colleagues locally and around the world. What is missing at this point is an integrated information architecture that consolidates the workspace, content, and tools that are increasingly included as a routine part of the educational and scholarly mission of the University.

Fedora Commons will collaborate with UVa on its new project to address this problem. The development of the Fedora-based Academic Information Space (AIS) will make it possible for scholars to easily create and re-purpose digital resources, rapidly moving them into classroom and research projects, while ensuring that users maintain control over their data for as long as they desire. Digital objects created in the AIS will be ready to move into a digital library if appropriate, without interruption from the community's point of view. The AIS will be an integrated software framework backed by Fedora Commons open source software. *A letter of support from Karin Wittenborg, University Librarian of University of Virginia, is attached to this proposal.*

Fedora Commons will also form a strategic partnership with Tufts University which is significant both from the perspective of digital libraries and from the standpoint of being on the cutting edge of humanities computing. Fedora is an enabling technology for the Perseus Project[27], a densely hyperlinked digital library of materials on Greek, Roman, and Renaissance culture, and one of the premier humanities computing projects. Dr. Greg Crane, founder and editor in chief of Perseus, is interested in the extent to which the World Wide Web not only enhances the work of researchers and students, but creates new audiences for cultural materials outside academia. Dr. Crane is exploring how computational humanities can help to democratize information without compromising intellectual rigor, and states that "our goal is not only to help traditional scholars conduct their research more effectively but, more importantly, to help humanists use the technology to redefine the relationship

[27] http://www.perseus.tufts.edu/

between their work and the broader intellectual community." [9] Dr. Crane was among the earliest adopters of Fedora and understands how Fedora Commons technologies can enable fundamental change in scholarship by facilitating new ways to relate and analyze information. We will work with Dr. Crane to evolve the Fedora Commons software to continue to support innovation in the humanities and cultural sectors. *A letter of support from Dr. Crane is attached to this proposal*.

In terms of new opportunities, Fedora Commons will leverage its partnership with NSDL to form new strategic partnerships with museums. To date, the Fedora Project has not had direct relationships with major museums, yet Fedora is being used successfully by many special collections libraries, historical societies, and cultural institutions to integrate museum-like objects into innovative digital libraries and learning environments. There is great potential for synergy between museums and the other communities already working with Fedora. With a focused outreach effort by Fedora Commons, we believe new and interesting museum partnerships can be formed. Our first outreach efforts will be with the Exploratorium and the American Museum of Natural History, both NSDL collaborators. Reflecting back on Section 2 and the discussion of the power of open source software, it is interesting to note that many museums have built their own systems or have invested in commercial technologies to manage and present their digital collections. There is great opportunity for Fedora Commons to provide a context for museums to become part of the new "open world" phenomenon.

6 Technical Strategy: Foster Innovation to Enable Revolutionary Change

Fedora Commons is uniquely positioned to integrate key information management paradigms – the Web, content management, and digital preservation – that are core components of a new information environment. The Fedora Commons will develop and deploy "net-centric" repository–based systems that will combine the richness of the Semantic Web, the interactivity of Web 2.0, and the reliability and quality-of-service of enterprise-grade system for managing the full lifecycle of digital information. As noted in the strategic technology forecast by Mills Davis (unveiled at the 2006 Semantic Technology Conference[28]), there is a challenge in building robust, self-protecting systems that simultaneously fulfill the integration and interoperability requirements of collaborative, knowledge-based communities:

> This vision of net-centric computing is often expressed as a "grand challenge" for the industry to develop systems and processes that are self declaring, self-integrating, self-optimizing, self-protecting and self-healing and that can scale from point-to-point semantic web services to pervasive service grids…. Large un-met needs exist at the "hinges of the business" — integration and interoperability needs that go across systems, across businesses, and across communities of interest. [11]

Fedora is already well positioned in that it integrates key technologies to that support this vision – repository, semantic, enterprise, and preservation technologies. However, an evolutionary system (vs. a static system) must be flexible and adaptable to ensure that the system remains relevant to users as technologies and behaviors change over time. Our vision is that Fedora Commons will provide an

[28] http://www.semantic-conference.com/

incubation and experimentation environment for new open-source components and services that fit the Fedora Commons vision, and will develop an appropriate technical framework that allows for multiple flavors of Fedora. Fedora Commons will provide leadership to ensure that these technical innovations integrate into the core software offering and reflect the quality that has always been part of the Fedora software distribution. A key task is to rationalize the architecture in a manner that permits flexibility. This could be achieved by evolving Fedora into a "plug-in" architecture (similar to the Eclipse framework). Alternatively, the Fedora Commons platform can support more configuration options. Another approach is to distribute different "solution bundles" which are pre-packaged configurations of Fedora tailored to particular use cases.

There are four particular areas of innovation that we will nurture in Fedora Commons. These are described in the next sections.

6.1 Full Leveraging of Semantic Technologies for building information networks

There is significant potential for semantic technologies to break through the barriers to interoperability for both content and services. To build truly distributed information networks that cross community and organizational boundaries, we need to move beyond current systems paradigms that often depend on strict schemas and the exchange of pre-defined structured metadata formats. "A semantics-enabled application infrastructure supports diverse content integration in an application setting by promoting storage of and operation on the meaning and relationships of data, rather than just its structure." [1]

In Fedora Commons, we recognize the potential of emerging semantic technologies for enabling content discovery and for accessing and understanding the meaning of information across organizational and system boundaries. Also, semantic technologies are emerging that can facilitate the discovery, composition, and orchestration of services that enhance and augment content (e.g., OWL-S [20] and related work).

As described in section 3.2, the current Fedora software enables semantic assertions within digital objects (as RDF), which are then indexed in a triple store to provide a view of an entire repository as a semantic graph that can be queried. This work was motivated by the Fedora Project's recognition that digital repository systems needed to be adapted to expose information as a semantic graph. Recent work by the Topaz project, has taken this notion one step further by demonstrating how a repository can be more natively graph-oriented (as opposed to reflecting a graph view of the repository). By collaborating both with the Fedora Project and the Mulgara Project[29], Topaz is developing a variation of the existing Fedora architecture where digital objects are modeled as patterns in a graph. Accordingly, digital objects are created and updated directly as transactions upon a graph and persisted in a Mulgara triple store. Simultaneously, digital object content byte streams are stored in a Fedora repository. Topaz, Fedora, and Mulgara are investigating what it means to build a repository architecture that more fully utilizes a graph store at its core.

[29] http://www.mulgara.org/

Another challenge is to federate semantic knowledge across multiple repositories. This is illustrated in Figure 6, which depicts a notional view of a graph integration and interoperability layer over distributed repositories. The Mulgara Project has already made progress in the area of distributed graph query that we will evaluate. There are other approaches that we will also consider. One method that is being investigated by our colleagues at Cornell and Los Alamos in the Open Archives Initiative Object Reuse and Exchange project[30] is a protocol and model-based approached for the exchange of named sub-graphs. Another method being investigated in the peer-to-peer community is the use of Distributed Hash Tables [14] to process complex RDF queries over distributed systems.

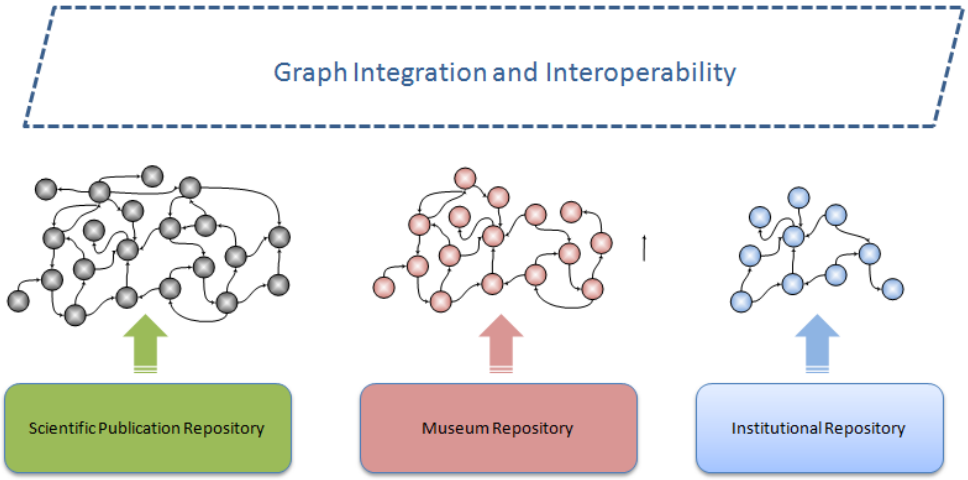

Figure 6 - Integrating semantic information from multiple repositories

Any solution we develop for integrating rich semantic networks into Fedora needs to co-exist with the other concerns of the platform: the notion of independent digital objects that are self-describing, portable, and "archive-ready"; enterprise orientation, with integration of workflow and messaging; and preservation orientation, incorporating monitoring, migration, and disaster tolerance. Optimally the Fedora system should be configurable to allow individual adopters to optimize for their particular applications and use cases.

6.2 Optimal integration of Collaborative "Web 2.0" Applications with Repositories

Over the past several years there has been a proliferation of tools and applications on the web that facilitate collaborative content creation, annotation and discussion, classification, and information sharing. Wikipedia[31] is certainly one of the best known of these applications. Another, in the area of bookmarking and tagging, is Connotea[32], developed by the Nature Publishing Group, which allows

[30] http://openarchives.org/ore
[31] http://en.wikipedia.org
[32] http://www.connotea.org

scientists to share and annotate references to the scientific literature. Zotero[33] provides a browser plug-in and tools to help scholars gather, organize, and analyze primary and secondary sources, automatically generate citations, and, ultimately shared this information with others. These applications are typically implemented as standalone applications and the relationships among information resources that originate from different systems are created through "mash-ups" or through simple hyper-linking within web pages.

Information that is created or gathered via collaborative web-based applications is typically stored within databases and other backend storage schemes that are particular to each application. There is not a common information infrastructure into which multiple such applications can create and access diverse content in an interoperable way. Nor is there any common approach to modeling information resources in a manner that promotes re-use in other contexts, or that facilitates a general-purpose management and preservation strategy for content.

Especially vulnerable is the content that is collected around the periphery of primary information resources. The fruits of the collaborative effort are typically in the form of asserted relationships among primary resources (e.g., articles, images, books, etc.), annotations of items, and the creation of metadata that augments or contextualized primary resources. This information is one of the main attractions of Web 2.0 applications, but without some thinking about the underlying storage and management architecture for this information, we risk losing it. The information that is created as a result of social interaction around primary information resources must be treated with the same respect as the primary materials themselves.

These observations motivate the need for a technical framework that enables Web 2.0 applications to be easily built upon a robust repository-based platform. The Fedora open source platform is already positioned to help. It provides storage and management for any type of content, including secondary information and relationships that are captured via Web 2.0 applications. All of its functionality is available via well-defined programmatic interfaces. Examples of Fedora as the underpinning for Web 2.0 type applications are already provided by existing Fedora-based projects.

Two notable applications are ExpertVoices[34] and OurNSDL, which are collaborative applications build by the NSDL that are integrated with its Fedora-based repository environment. ExpertVoices is a blogging environment that enables experts to comment on scientific resources in the NSDL digital library and to relate those resources to real-world science events. ExpertVoices provides the infrastructure for engaging teachers, scientists, librarians, and students in conversations about science topics. It is built using a standard, open-source blogging system (WordPress MultiUser[35]), with a plug-in that was developed to store the blogs - and their relationships to NSDL resources - in the Fedora-based NSDL repository. Similarly, OurNSDL builds on wiki technology to allow communities of users to create, annotate, and organize NSDL resources. Again, leveraging open source on all fronts, the NSDL is

[33] http://www.zotero.org

[34] http://expertvoices.nsdl.org/

[35] http://mu.wordpress.org/

creating a plug-in for the popular MediaWiki[36] software to store wiki text and relationships within the Fedora-based NSDL repository.

The Topaz project is pursuing another strategy in its integration of the PLoS ONE collaborative web-based tools with its repository-based platform built upon Fedora and Mulgara. In this work, Topaz has taken up the challenge of capturing fine-grained annotations of articles – meaning that annotations are associated with fragments of articles, or even particular words. Each annotation must be access-controlled so that only the author of the annotation can made changes to it. When users of PLoS ONE retrieve an article, all annotations associated with that article are also retrieved for quick viewing. To accomplish this, a complex graph of information must be quickly and efficiently retrieved and presented to the user. This graph is constantly subject to change as new and existing users add and modify annotations. Topaz has leveraged the advanced capabilities of the Mulgara triplestore to provide a rich graph-oriented interface to objects and their annotations, with a Fedora repository providing persistent storage for article byte streams.

While these applications demonstrate integration of collaborative applications with repository-based management environments, a number of key challenges remain that Fedora Commons will continue to work on. We are discussing with Topaz developers strategies for improved integration of Mulgara and Fedora to ensure that annotations stored in the Mulgara triplestore and articles in Fedora are both treated as first class information entities that can be managed in a consistent manner, amenable to long-term preservation. Also, a number of Fedora collaborators have identified the need for improved interfaces for operating on networks of related objects in Fedora repositories. While Fedora currently supports the creation and access of digital objects and the assertion of relationships among them, its current API is optimized to operate on one object at a time, not to operate on a graph of related objects as the target unit of management for create/read/update/delete operations. We are already exploring alternative ways to modify the existing Fedora implementation to better support this notion. These are examples of areas in which Fedora Commons can foster innovation to make it easier for developers of collaborative, dynamic applications to easily and effectively integrate with a robust repository core.

6.3 Workflow Integration with Repositories and Enterprise-Orientation

Starting in 2006, Fedora community collaborators participated in the Fedora Project's Workflow Working Group to discuss the use cases that motivate building a workflow service into the Fedora open-source system. The group concluded that a configurable workflow service integrated with repository services is a key requirement for creating robust collaborative scholarly and scientific collaboration and communication systems. Especially important requirements are the tracking of human tasks and the orchestration of automated services to achieve a particular outcome such as: (1) the creation of an intellectual work by multiple authors, (2) the ingestion and validation of complex content into a repository, especially in creating "super digital objects" comprised of a set of related digital objects, (3)

[36] http://www.mediawiki.org

the publication of new forms of scholarship, and (4) the execution of preservation workflows to manage and protect digital scholarship after it has been created.

With support from the Andrew W. Mellon Foundation (2007-2009), the existing Fedora Project has already begun to explore the potential and challenges of integrating workflow capabilities into the Fedora platform. Over the course of the next two years, this work will focus on evolving the Fedora system to directly support, and be compatible with, enterprise system features that include workflow engines, messaging brokers, and distributed transaction managers. The requirement is that all of these integrations will be in the form of open-source software that can be distributed with Fedora. The results of the Mellon-funded effort will be a key contribution that will lead the way in evolving the Fedora Commons technical platform.

6.4 Archiving and Digital Preservation

Major institutions have become aware of the need to provide custodial care and long-term access to digital content. Examples include the Library of Congress and the National Archive and Records Administration (NARA) in the U.S., as well as the Joint Information Systems Committee in the U.K. Institutional funding has been provided by philanthropic organizations such as the Mellon Foundation and Institute of Museum and Library Services (IMLS)[37] to develop techniques and software that can be used by institutions to enable their digital preservation missions. A number of universities and cultural institutions have begun digital preservation projects such as the Yale/Tufts contributions to records management and the development of preservation services at Rutgers University. In the scientific community, National Oceanographic and Atmospheric Administration (NOAA) has recognized the value of long term climate data in developing the requirements for the CLASS archive. The San Diego Super Computing Center has been a major player in digital preservation, and is the creator of the well-known Storage Resource Broker (SRB) technology[38]. Finally, the Research Libraries Group (RLG) and NARA have published a checklist and methodology to support the certification of "trusted digital repositories."

Despite these efforts, there is currently no open source software framework that integrates all of the building blocks for constructing a trusted digital repository capable of long term digital preservation. Few commercial or open-source content creation and management systems incorporate *any* digital preservation capabilities. Actually, many of them interfere with the addition of such features by keeping key data elements inaccessible as proprietary features of their products. Or these products cannot be incorporated easily as components of an integrated digital repository environment.

Our work with Fedora as a preservation framework extends back to the early year of the Fedora project [24]. The existing Fedora open source system was designed to provide a cohesive, well-informed, over-arching, enterprise systems architecture for implementation of trusted digital repositories. The original Fedora repository service was designed from the beginning to enable digital preservation strategies. Fedora implements the design principles of a self-describing, self-healing system in several ways. First,

[37] http://www.imls.gov/
[38] http://www.sdsc.edu/srb/index.php/Main_Page

in Fedora, content (and metadata) is encapsulated in self-defining digital objects and then indexed into related digital object graphs. The repository's digital objects and their relationships can be stored as files and later recovered, and the graph index can be rebuilt, even if the repository software no longer exists. Another interesting feature of Fedora is that digital objects can be self-delivering by use of the Fedora disseminator feature. For example, a "disseminator" can be used to associate a behavior contract and supporting services with digital objects to provide transformations and views of obsolete content formats. Fedora also has the potential to act as a registry repository that can store the identity of content formats with any desired precision using external authorities such as PRONOM[39] and the Global Digital Format Registry[40]. Fedora can even be used to archive software associated with the digital content for future use. Finally, as a service-oriented architecture, Fedora permits the integration of third-party preservation service components within the Fedora system.

Our attention to preservation as an integral part of the Fedora architecture has been recognized by a number of preservation-focused projects and experts in the field. Recently a Fedora-based project from Harris Corporation was one of the two finalists for a bid from NARA to build the Electronic Records Archives (ERA) system [2]. Harris developed a prototype that demonstrated how the preservation-aware features of the Fedora architecture could be leveraged for a large-scale high-demand application. Although NARA selected Lockheed Martin to build the ERA system, they remain interested in the Harris open-source proposal due to its key insights about the significant of open systems and open software to the preservation problem. In another part of this proposal we describe the work of another project by Johns Hopkins University and the National Virtual Observatory that used Fedora in its prototype of a "preservation appliance" for data curation and archiving [12]. Finally, the Fedora Preservation Services Working Group[41] is a growing community of experts from libraries and archives that is developing a general definition of preservation services for the Fedora service framework. The charge of the group is to recommend enhancements to the Fedora repository service as well develop specifications for new preservation-support services for the Fedora Service Framework.

Digital preservation and archiving is widely recognized by experts to be a multi-faceted problem with no single solution [13]. It is a problem that requires flexible technology, community participation, and organizational commitment. As described in this section, Fedora Commons begins with these foundations in place: (1) the technology is preservation-aware and constructed to allow application-specific plug-ins, (2) a Fedora community-based working group is already in action, and (3) the partnerships described in this proposal demonstrate initial organizational commitment. Our plan is to leverage all these factors and make Fedora Commons a context for innovative solutions to digital preservation.

[39] http://www.nationalarchives.gov.uk/pronom/
[40] http://hul.harvard.edu/gdfr
[41] http://www.fedora.info/wiki/index.php/Working_Group:_Preservation

7 Fedora Commons Budget Summary

This proposal requests **$4.3 million from the Moore Foundation** to support the startup phase of Fedora Commons. This is a portion of a forecasted total budget of $5.6 million to achieve the technical and community development goals of Fedora Commons over the course of the four year startup period. Other funds will be secured to meet the difference between the total forecasted budget and the funds requested from Moore. We have already received additional funding for Years 1 and 2 in the form of a $518,000 grant to Cornell University from the Andrew W. Mellon Foundation to support the integration of a workflow engine and to improve the enterprise orientation of the Fedora system. In addition to this, we anticipate that the Fedora Commons outreach effort will enable us to secure additional funds in Years 3 and 4 to meet the forecasted total budget. We forecast additional income of $250,000 in Year 3 and $550,000 in Year 4. The anticipated sources of this income will be a combination of grants from other foundations, gifts/donations, membership fees, and consulting income, all described below.

The four year startup budget is detailed in Appendix A. The most significant thing to note is the proposed manner in which the funds will be distributed to achieve a smooth transition of the existing Fedora Project (currently at Cornell University and University of Virginia) to the new Fedora Commons non-profit organization. We propose a solution that leverages the commitment that Cornell University has to Fedora and its future. Currently, the existing Fedora leadership and developers are grant-funded employees of the Cornell Information Science research and development group. We therefore propose that Fedora Commons execute a subcontract in Years 1 and 2 with Cornell University to fund the existing Fedora team (whose current Mellon funding expires in September 2007). This will have benefits to both Fedora Commons and Cornell University. It will streamline the Fedora Commons startup phase by leveraging the existing business and human resource processes of Cornell University. It will provide an interim home office for Fedora Commons - both a legal address and office space. It will provide time to plan and undertake the transition of intellectual property and to startup the necessary businesses processes in the new non-profit to support a fully-staffed organization. As the institution where Fedora was originally invented, Cornell has generously agreed to this transition plan recognizing the benefits of association with this enterprise.

A similar transition model has already been executed by other university-based open-source projects that have moved into independent non-profits corporations. For example, the Sakai Foundation[42] funds its staff via subcontracts to the University of Michigan and other member universities. This enables university faculty and staff to maintain their existing positions while directly working on the deliverables to meet the goals of the Sakai Foundation. The Kuali Foundation[43] follows the same pattern, with relationships spanning many universities, including Indiana University and Cornell University. With appropriate contracts and Memorandums of Understanding in place, this model has proven effective in the initial transition period from a university-based project to a non-profit organization.

[42] http://sakaiproject.org
[43] http://www.kuali.org/

We propose a slightly different model from Sakai and Kuali in which the critical leadership roles of Fedora Commons are either full-time or part-time employees of the Fedora Commons non-profit corporation at initial startup. To ensure fiscal responsibility for the grant funds, and proper dispersal of funds, we propose that a Chief Financial Officer be hired immediately and directly into the non-profit organization. Additionally, we propose that the role of the Executive Director straddle Fedora Commons and Cornell University during the first two years (50/50 FTE in each organization). The existing Fedora Director at Cornell, Sandy Payette, will fulfill the dual role of Executive Director of Fedora Commons, and Director and Principal Investigator of Fedora at Cornell. As Executive Director, she will be in charge of the vision and operation of Fedora Commons, while simultaneously directing the subcontract team working for Fedora Commons at Cornell. This leadership model provides complete integration of Fedora Commons with the Cornell subcontract under which the bulk of the work will be done in the first two years of the startup phase.

Starting in Year 3, all employees will transition to the Fedora Commons organization. At this point the non-profit corporation will be stable and in a position to support a full organization. The budget detail in Appendix A shows that starting in Year 3, the Executive Director will become a full-time employee of the Fedora Commons organization, and all other key roles will be direct employees of Fedora Commons. The final two years of the startup phase will operate with this organization, which will eventually become the organizational model for the sustainability phase of Fedora Commons. Appendix B details the projected budget and income for the sustainability phase (which begins in 2011). The next section describes the strategy for maintaining the sustainability phase.

8 Fedora Commons Sustainability and Business Plan

The sustainability of Fedora Commons will be ensured by a combination of committed organizations providing "in-kind" contributions (e.g., development resources) and the securing of new sources of *tangible* income. We will take a multi-faceted strategy to sustainability.

8.1 Community commitment and in-kind contribution of resources

The sustainability plan for Fedora Commons is built on the assumption of a thriving and committed community. Commitment can be expressed in many ways, including the giving of gifts or paying of fees (described in the sections below), but also as the giving of time and effort, known as "in-kind" contribution. This form of giving is usually in the form of an individual or an organization devoting human resources to work on an intellectual activity, design/development of software, or other tasks that forward the organizational mission.

These in-kind contributions are a key component of the Fedora Commons business plan. Community participation will enable Fedora Commons to operate with a lean core team. As described in Section 8, by the end of the startup phase of Fedora Common, most programming and system development will be done by community "committers," coordinated by the Fedora Commons chief architect. Community members will also take on a significant role in the defining of new requirements for the Fedora

Commons system. The Fedora Commons outreach effort during its startup years will ensure that a critical mass of individuals and institutions commit to providing human resources to Fedora Commons.

As discussed in Section 2, there is a recognized economic benefit to community participation in the development of open source software. The Fedora Commons "in-kind" strategy is validated by the commons-based peer production model that has been described by Benkler and others.

8.2 Industry Sponsorship and Gifts

As part of the Fedora Commons outreach effort, we will develop relationships with commercial organizations that have a strategic benefit in aligning with Fedora Commons. This is an important sustainability strategy taken up by most successful open source projects.

Already, the news of the formation of Fedora Commons has already spurred interest by potential industry players. For example, Sun Microsystems has recently approached the Fedora directors to discuss the prospect of forming a relationship in the context of synergies between Fedora repositories and Sun's Honeycomb archival storage system. We plan to leverage existing relationships with commercial users of Fedora (e.g., VTLS, Harris, Fiz Karlsruhe, RightsCom). Beyond this, we will focus on developing new relationships with industry players that have already expressed an interest in Fedora (e.g., Microsoft), and we will identify and pursue fertile prospects (for example, companies focused on semantic technologies, storage technologies, database systems, content management, strategic business information, library and archive systems vendors, and others markets).

8.3 Consulting and Services for Fee

The existing Fedora Project has already observed that there is a market for consulting services around the use of open-source technologies to build sophisticated knowledge-based systems, digital archives, and enterprise systems in support scholarship and science. Members of the existing Fedora team have already consulted to many organizations, and we expect that this demand can be turned into an income-generating service that can be operated by Fedora Commons.

The Fedora Commons software is extremely flexible, but with this flexibility comes decisions on how to best use it to achieve the goals of particular use cases. As with any systems building endeavor, there are challenges of information modeling and systems integration that may require expertise that does not exist within an organization.

In the case of building systems and applications upon the Fedora Commons platform, there are interesting choices to be made by institutions in terms of how to model their digital objects and the relationships among them (often referred to as the "content modeling "challenge in the Fedora community). There are questions on how to build workflows to ingest content, or to help preserve content. There are questions of how to use different interoperability standards with Fedora. There are choices to be made in how services will be integrated to best enable the applications built on top of the Fedora Commons platform.

We plan to create a business plan for consulting service around these types of questions. Consulting income can be directly reinvested in the Fedora Commons organization to help sustain it. The Executive Director and Business Strategist of Fedora Commons will work together in the startup years to determine the best model for offering such services.

8.4 Fedora Commons Partners Program

Earlier, we introduced the notion of strategic partners as organizations who will work closely with Fedora Commons in developing flagship systems, defining new system requirements, and devising strategies for outreach and community development in the target sectors. These initial strategic partners will not be voting members of the Fedora Commons organization. Instead, they will be collaborators that contribute "in kind" to Fedora Commons.

Over the course of the startup period of Fedora Commons, however, it will be important to define an appropriate status for organizations that we deem "partners" beyond the scope of the initial group of strategic partners. There are several possible models for this. The Fedora Commons organization can have either voting members or non-voting members. A key motivation in this decision is the collection of membership fees from committed organization that have a long term interest in Fedora Commons.

Membership fees will be an important source of income for Fedora Commons. While it is not necessary to create a membership organization to collect fees, it may be desirable. The exact nature of membership and the fee model will be determined by the Fedora Commons Board of Directors at an appropriate time during the four-year startup period.

Whether partners of Fedora Commons are voting members of the corporation, or not, we anticipate that organizations will be motivated to pay a fee that will entitle them to a suite of benefits. We envision these benefits to include: (1) participation in advisory councils, (2) participation in architecture councils, (3) admission to all events and conferences sponsored by Fedora Commons, and (4) voting on the prioritization of development items in Fedora Commons roadmap.

9 Fedora Commons Governance

The Fedora Commons organization is being incorporated in the State of New York, and will be a 501c3 non-profit corporation. The initial organizational structure of Fedora Commons is depicted in Figure 7. The Board of Directors will have complete and plenary power to manage, control, and conduct all the affairs of the corporation. The initial members of the Board of Directors are listed in Appendix C. The Board will have the power to appoint new Board members and will appoint the Executive Director (who will be the Chief Executive Officer of the non-profit corporation). Fedora Commons is not initially incorporated as a voting membership organization. This is intentional to give the Fedora Commons Board of Directors full control of the corporation in the start-up years. However, the Board can decide at any time to change the Fedora Commons articles and bylaws to introduce the possibility of voting members of the non-profit corporation. Many open source organizations are incorporated with such a membership model, for example Sakai, Eclipse, Kuali are all voting member organizations.

It is expected that the Board will immediately appoint Sandra Payette as the Executive Director. Payette is the Incorporator of Fedora Commons and the current Director of the Fedora Project at Cornell University. The Board will also appoint an interim Chief Financial Officer (CFO) who will take fiscal responsibility during the startup of Fedora Commons and manage the grant funds received from the Moore Foundation. The interim CFO will step down when the permanent position of "Chief Financial Officer and Business Strategist" is filled. As, depicted in the organizational chart below, the Executive Director reports to the Board of Directors and the CFO/Business Strategist reports to the Executive Director. Together these two officers will evolve the business strategy for Fedora Commons and implement this strategy to ensure sustainability beyond the initial four year startup funding period.

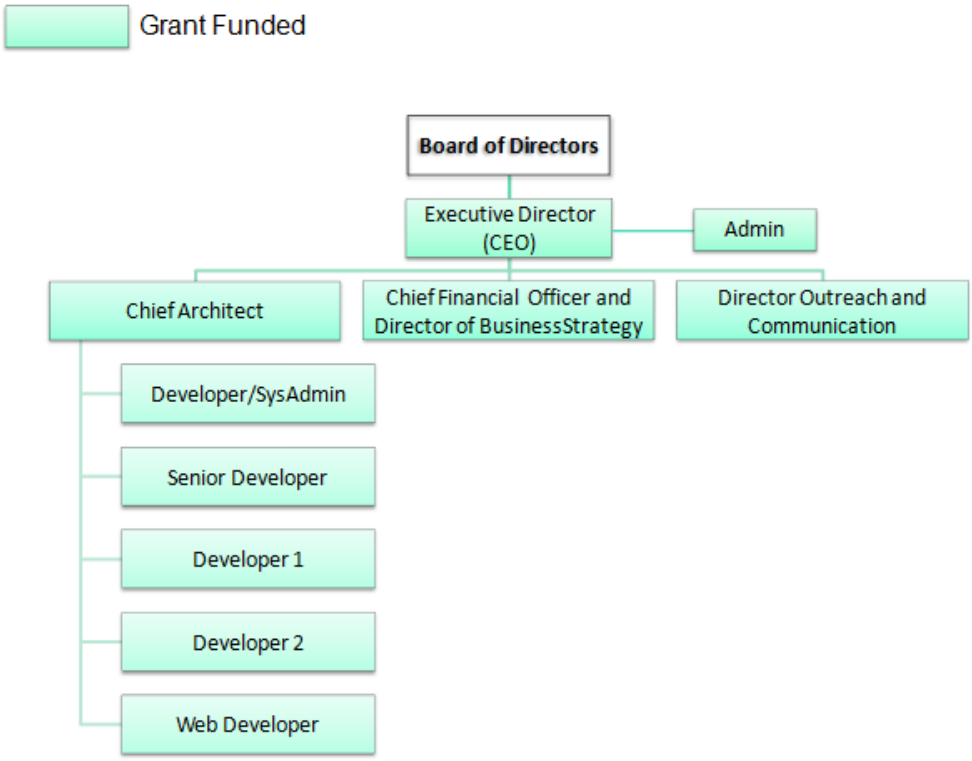

Figure 7 - Fedora Commons – Start-Up (2007)

Other key roles that will report to the Executive Director are the Chief Architect and the Director of Outreach and Communication. Together the Executive Director and the Chief Architect will work together to ensure that the technical vision for the Fedora Commons is successfully implemented. Reporting to the Chief Architect are the technical developers required to focus on the evolution and maintenance of the Fedora Commons software (described in Section 6). The Chief Architect will also be responsible for coordinating community software development activity. In support of community-building, the Executive Director will also work with the Director of Outreach and Communication to develop strategies for propagating the Fedora Commons vision, building awareness of Fedora Commons, and increasing the number of institutions that use Fedora Common software. Over the course of the startup phase, the Director of Outreach and Communication will establish community-

based councils that will focus on defining and prioritizing new system requirements for the four target community sectors and also participate in the outreach and community-building efforts of Fedora Commons.

By the end of the four year startup period, the Fedora Commons organization will have transitioned to the organizational structure necessary to support long-term sustainability. At this point, Fedora Commons will have secured funds and generated income sufficient to meet the projected sustainability budget for the year 2011-2012, which is projected to be $1.5 million.

Figure 8 depicts what the Fedora Commons organization will look like at the beginning of the sustainability phase. At this point, Fedora Commons will continue to employ the core leadership team (Executive Director/CEO, CFO/Business Strategist, Chief Architect, Director of Outreach), plus one system administrator, one senior developer, and an administrative assistant. As discussed in Section 2, strong leadership is essential to successful open source projects, and the proposed organizational structure is consistent with that of other successful open source projects in the academic and scholarly middleware and application space (e.g., Kuali and Sakai) and in the broader technical infrastructure space (e.g., Eclipse and Apache).

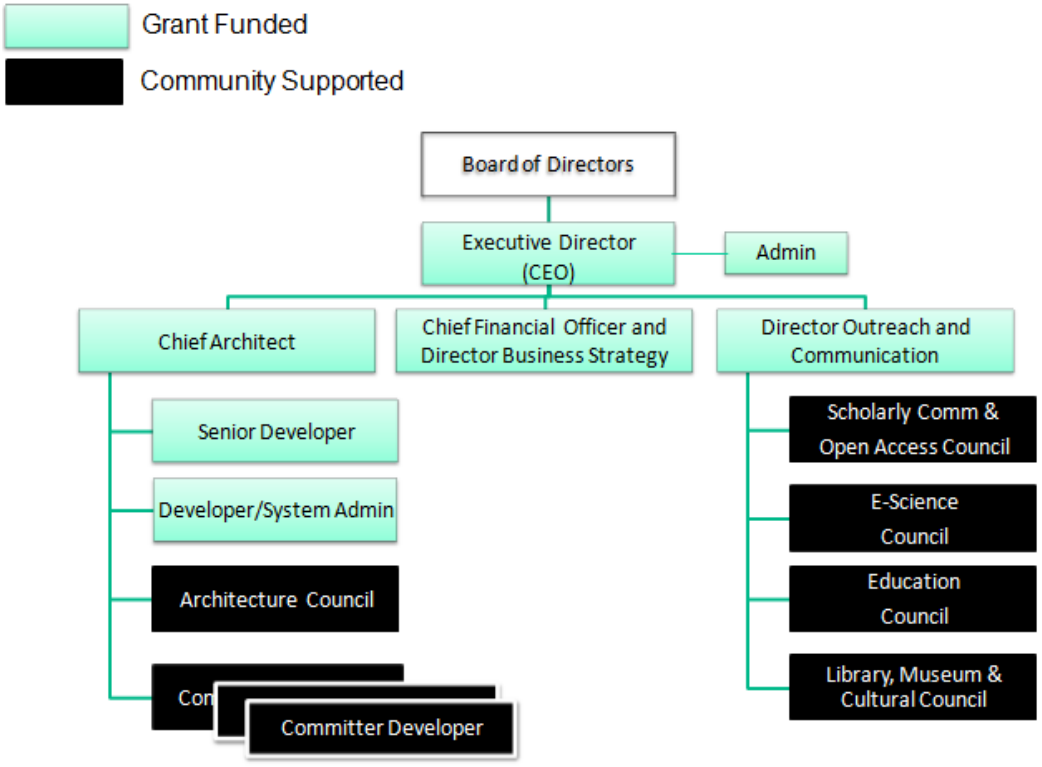

Figure 8 - Fedora Commons – Sustainability Stage (2011-onward)

As discussed in Section 2, a key characteristic of successful open source projects is the development of strong community. As part of the community outreach effort, the Chief Architect will establish and lead a community-based Architecture Council which will be the coordinating entity for collaborative

community system design and software development. The Architecture Council will also be responsible for the Fedora Commons software roadmap. A group of new community developers ("committers") will take on the majority of new software development, as well as maintenance of the existing software, under the direction of the Architecture Council and the Chief Architect. The community will also participate in setting future directions and defining new requirements via Sector Councils – one council for each target community sector. The community councils and committers are depicted in Figure 8.

In the steady state organizational model, it is still possible for Fedora Commons to directly employ additional staff. This could be done if the organization received gifts in excess of the baseline budget, or if an organization makes a contribution of funds to be targeted at hiring developer(s) to focus on particular pieces of work. This is common in other large open source organizations such as Apache and Eclipse.

10 Summary of Grant Outcomes and Outputs

The overall outcome for the Fedora Commons proposal is to enable the organizational and technical frameworks necessary for sustainable open-source software to support revolutionary change in how scientists, scholars, and educators produce and share their intellectual outputs, and ensure the integrity and longevity of information.

A major measure of the success of this outcome will be to increase the number of institutions that commit to sustaining the Fedora Commons organization by 100% per year for 4 years. This will be done by outreach to the Fedora community with the goal of increasing the numbers of institutions that commit to Fedora Commons. An institution can demonstrate commitment to Fedora Commons by providing human resources (architects, programmers, consultants, and advisors) to evolve and maintain the Fedora Commons system. Another measure of success is the securing of other tangible funds as income to Fedora Commons. This can be achieved by organizations becoming fee-paying partners of Fedora Commons, as well as through the securing of monetary donations and gifts to the Fedora Commons organization from both commercial and non-commercial organizations. Success can be measured by securing funds equal to or greater than the projected "steady state" budget which begins at the end of June 2011.

Specific Outputs of this grant include:

- Deliver the next generation of the Fedora repository system, to be known as the Fedora Commons open platform.
- Integrate essential new open-source "enterprise" components of the Fedora Commons open source platform including a workflow engine and messaging broker.
- Develop new services for the Fedora Commons open source platform including digital preservation services and adapters to integrate information across heterogeneous systems.

- Demonstrate ***five*** mature "flagship" Fedora-based collaborative systems in each of ***four*** target sectors, and measure system integrity and sustainability with appropriate community standards (e.g., NARA[44], ISO/OAIS[45], etc.)
- Increase the number of installations of the Fedora Commons system worldwide
- Generate new sources of income to meet or exceed the projected budget for the sustainability phase (see Appendix B)

[44] National Archives and Record Administration (NARA) audit checklist for trusted repositories.
[45] ISO Standard for Open Archival Information System (OAIS)

References

1. IBM Semantic Layered Research Platform: Towards a Semantics-enabled Application Infrastructure, IBM Advanced Technology Group, Cambridge, MA, 2006.
2. National Archives Names Lockheed Martin to Build Archives of the Future, The National Archives, 2005.
3. Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven, CT, 2006.
4. Berlind, D. Open Source: IBM's deadly weapon *ZDNet*, 2002.
5. Bishop, A.P. and Star, S.L. Social informatics of digital library use and infrastructure. in Williams, M.E. ed. *Annual review of information science and technology: Vol. 31*, Information Today, Medford, NJ, 1996, 301-401.
6. Carlson, S. Lost in a Sea of Science Data *Chronicle of Higher Education*, 2006.
7. Choudhury, S. and Martino, J. A Technology Analysis of Repository Services, *CNI Task Force Meeting*, Washington DC, 2005.
8. Coleman, G. The Social Construction of Freedom in Free and Open Source Software: Hackers, Ethics, and the Liberal Tradition *Anthropology*, University Of Chicago, Chicago, 2005.
9. Crane, G. The Perseeus Project and Beyond. *D-Lib Magazine*, *January 1998*.
10. David, S. The Open World *Science, Technology, and Society*, Cornell University, Ithaca, 2007.
11. Davis, M. Semantic Wave 2006. 10X, Project 10X, Washington, DC, 2006.
12. Hanisch, R., Choudhury, S., DiLauro, T., Szalay, A., Vishniack, E., Steffen, J., Ehling, T., Milkey, R. and Plante, R. Digital Data Preservation in Astronomy: A Collaboration Among Libraries, Publishers, and the Virtual Observatory *2nd International Digital Curation Conference*, Glasgow, 2006.
13. Hedstrom, M. and Ross, S. Preservation research and sustainable digital libraries. *International Journal on Digital Libraries*, *5* (4). 317-324.
14. Heine, F., Hovestadt, M. and Kao, O. Processing complex RDF queries over P2P networks *2005 ACM workshop on Information retrieval in peer-to-peer networks* ACM, Bremen, Germany, 2005.
15. Hoe, N.S. Breaking Barriers. The Potential for Free and Open Source Software for Sustainable Human Development, United Nations Development Program, 2006.
16. Kelty, C.M. *Two Bits: The cultural significance of Free Software and the Internet*. MIT Press, Cambridge, MA, 2006.
17. Lassila, O. and Swick, R.R. Resource Description Framework: (RDF) Model and Syntax Specification, World Wide Web Consortium, 1999.
18. Lessig, L. *Free culture : how big media uses technology and the law to lock down culture and control creativity*. Penguin Press, New York, 2004.
19. Lessig, L. *The future of ideas : the fate of the commons in a connected world*. Random House, New York, 2001.
20. Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., Sirin, E., Srinivasan, N. and Sycara, K. OWL-S: Semantic Markup for Web Services, W3C, 2004.
21. National Science Foundation Cyberinfrastructure Panel. Cyberinfrastructure Vision for 21st Century Discovery, National Science Foundation, Washington, D.C., 2007.
22. O'Reilly, T. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, O'Reilly, 2004.

23. Payette, S. and Lagoze, C., Flexible and Extensible Digital Object and Repository Architecture. in *Research and Advanced Technology for Digital Libraries, Second European Conference, ECDL '98*, (Berlin, Heidelberg, 1998), Springer (Lecture notes in computer science).
24. Payette, S. and Lagoze, C. Value-Added Surrogates for Distributed Content: Establishing a Virtual Control Zone. *D-Lib Magazine*, *6* (6).
25. Sierra, K. and Russell, D. Creating Passionate Users.
26. Staples, T. and Wayland, R. Virginia Dons FEDORA: A Prototype for a Digital Object Repository. *D-Lib Magazine*, *July*.
27. Steinberg, D.H. Open Source and Web 2.0 *O'Reilly Network*, 2005.
28. Szulik, M.J. Open for Change. *EDUCAUSE Review*, *42* (1).
29. Treloar, A., Building an Institutional Research Repository from the Ground Up: The ARROW Experience. in *AusWeb04*, (Gold Coast, Australia, 2003).
30. Treloar, A. The Dataset Acquisition, Accessibility, and Annotation e-Research Technologies (DART) Project: building the new collaborative e-research infrastructure *AusWeb 2006*, Noosas Lakes, Australia, 2006.
31. University of Oxford. Building Preservation Environments *Oxford Internet Institute*, 2005.
32. Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C. and Warner, S. Rethinking Scholarly Communication: Building the System that Scholars Deserve. *D-Lib Magazine* (September 2005).
33. Vest, C.M. Open Content and the Emerging Global Meta-University. *EDUCAUSE Review*, *41* (3).
34. Warner, S., Bekaert, J., Lagoze, C., Liu, X., Payette, S. and Van de Sompel, H. Pathways: Augmenting interoperability across scholarly repositories. *International Journal on Digital Libraries special issue on Digital Libraries and eScience*, *forthcoming*.
35. Weber, S. *The success of open source*. Harvard University Press, Cambridge, Mass., 2004.
36. Wheeler, B. Open Source 2010: Reflections on 2007. *EDUCAUSE Review*, *42* (1).
37. Williams, R., Moore, R. and Hanisch, R. A Virtual Observatory Vision based on Publishing and Virtual Data, US National Virtual Observatory, 2003.
38. Zia, L.L. The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program. *D-Lib Magazine*, *8* (11).