

Vol. 1, Number 2, June 4, 2008

Read or download online at: <http://fedora-commons.org/resources/newsletter.php>

Archived issues are available at: <http://fedora-commons.org/resources/archive.php>

Subscribe to HatCheck RSS at: <http://expertvoices.nsd.org/hatchcheck/feed/>

Fedora Commons Overview

The Many Faces of Collaboration with Fedora Commons by *Sandy Payette*

Engineering Punchlist

Fedora 3.0 Release Schedule and Features by *Dan Davis*

News

- **Muradora 1.3 Release** by *Chi Nguyen*
- **Stargazing: Relating Information and Applications to the Data Workflow** by *Sayed Choudhury*
- **Bringing Seven Centuries into the Future: Bibliotheque Nationale de France** by *Thomas Ledoux*
- **Spoken Word Services Delivers Digital Audio and Visual Content** by *Iain Wallace*
- **Fedora Performance and Scalability Wiki Online** by *Matthias Razum*
- **Plug a Wiki into a Fedora Repository and Get... A Scholarly Publication** by *Carol Minton Morris*
- **Object Reuse and Exchange (OAI-ORE) Standards Beta Release** by *Carl Lagoze*
- **Road Report from PASIG** by *Carol Minton Morris*

Upcoming Events

- eSciDoc Days 2008 in Berlin
- Schedule of Summertime Fedora Commons Community Meetings and Events
- Of Repository Road Shows and Fringe Festivals
- Carnegie Foundation Open Education Webevent

Bookmarks

- Download the "Roadmap to Fedora Development"
- NSDL Director Kay Howe at JA-SIG, "Ubiquity, Interdependence, and the Age of Collaboration"
- Fedora Commons Quick Poll I Results: User Gatherings
- Strands of Global Web of Knowledge Come Together at Third International Open Repositories Conference 2008



In each issue of HatCheck, Director of Community Strategy and Outreach, Thornton Staples, will introduce a topic of interest from the Fedora Commons Community Registry. Contact us (<http://fedora-commons.org/contact.php?id=communityregistry>) to add your project to the Registry.

Fedora Commons is often asked, "Who is using Fedora?" The current answer is far from accurate. Fedora open repository platform software is downloaded approximately 2,000 times each month while the Fedora Commons Community Registry lists a community of only 124 projects, organizations and institutions. The potential pool of Fedora Commons community members is estimated to be many times bigger.

Where is everybody? The answer is up to you. If you haven't already seen the Fedora Commons Community Registry (FCCR) please take a look. Any project, organization or institution not listed that is actively using Fedora, or helping someone else use Fedora, is encouraged to "add to the list" by using the link at the top of the page to contribute FCCR information.

Sustaining Fedora Commons depends on the strength of the community that makes use of the Fedora open source repository platform.

(continued on page 12)

HatCheck is a quarterly newsletter published by Fedora Commons (<http://fedora-commons.org>) that is provided by and for communities of Fedora users, developers, vendors and decision makers to facilitate the exchange of news, events, information and achievements. Your comments and content are most welcome! Contact Carol Minton Morris, (communications@fedora-commons.org; 607 255-2702) Fedora Commons Director of Communication and Outreach, to contribute news and information for the September 4, 2008 issue of HatCheck.

Fedora Commons Overview

The Many Faces of Collaboration with Fedora Commons

By Sandy Payette, Executive Director, Fedora Commons

Ithaca, NY — In the last issue of *HatCheck*, I mentioned the importance of collaboration in achieving the Fedora Commons mission. Over the past months, I have had the opportunity to communicate with many of you to discuss your ideas for advancing our open source software to provide new capabilities for preserving and archiving digital content, fitting repositories behind collaborative web applications, enabling interoperability among scholarly and scientific repositories, and supporting the important task of data curation. With your input we have created the new Fedora Commons Technology Roadmap, which is now available online at <http://www.fedora-commons.org/resources/roadmap.php>.

As we plan ahead for the next 6-12 months, collaboration will be essential in achieving our goals. But we cannot take this assertion at face value. In moving forward with any collaboration, it is important to understand why we collaborate and what types of collaborations are possible, and to review the essential nature of successful collaborations. With so many potential collaborations available to a new non-profit such as Fedora Commons, it is important that we have a dialog with potential partners in which we ask basic questions such as what are the benefits of collaboration, in what manner might we collaborate, and how does the proposed collaboration integrate with other partnerships? We may want to work together in some manner, but may not be able to get to the next step of defining clear goals that ensure mutual benefit. In other cases, we may believe that a complex strategic planning process is necessary, and we may miss some low hanging fruit. It may be that we know what we would like to achieve if we could work together, but there may be difficulty in committing the necessary resources at the right time. Then there is the challenge of plenty: we could engage in so many collaborative opportunities that we could run the risk of diluting focus. How does any organization navigate this terrain?

For Fedora Commons, the answer has been to define three basic models of collaboration and to address each one with a different approach. The most basic form of collaboration is when an individual participates in the Fedora open source software development process by contributing bug fixes or developing new features. We have a simple process for becoming a contributor or a committer, and we encourage many people to join in. The first step is to contact Chris Wilper, Lead Developer, (cwilper@fedora-commons.org) or Dan Davis, Chief Software Architect (ddavis@fedora-commons.org).

The next form of collaboration is where a group of community members band together around a particular problem and work towards a collective contribution to Fedora Commons aimed at solving that problem. We have approached this type of collaboration model by adopting a community organizing approach. Under the leadership of Thorny Staples, we are helping the community organize “solution councils” to work together in defining

a particular problem, outlining goals, and devoting resources towards achieving those goals. Contributions can be in the form of requirements specifications, new software modules/services, or new applications that can be contributed back to Fedora Commons. Currently, we have four such solution councils under formation, each with a community champion (chairperson): the Preservation and Archiving council (Ron Jantz, Rutgers), the Data Curation council (Sayeed Choudhury, Johns Hopkins), the Open Access Publishing council (Rich Cave, PLoS), and the Applications Integration council (Matt Zumwalt, MediaShelf).

The most advanced type of collaboration is what we have referred to as “strategic partnerships.” This form of collaboration is characterized by Fedora Commons forming an alliance with another organization – either a company or another not-for-profit organization. This third type of collaboration requires contemplation and strategic planning to ensure success. I encourage anyone forming strategic alliances with other organizations to consider the partnership evaluation framework described by James Austin in his book *The Collaboration Challenge* (see citation at <http://www.worldcat.org/oclc/42934642>). Austin offers guidelines known as “The Seven C’s” for evaluating whether a partnership is likely to have key elements for success. Briefly, Austin provides a framework of questions to ask in making this determination:

- **Clarity:** What is the purpose of the collaboration? Has each partner determined the function and value of the collaboration and how it fits in the plan of achieving its goals?
- **Connection:** Do the individuals involved have good interpersonal connections? Do the leaders of the organizations have high quality interactions? Do the people believe in the collaboration?
- **Congruence:** Do the partners understand the mission, strategy, and values of each other? How can each partner help the other in accomplishing the mission?
- **Creation of value:** What specific benefits will accrue to each party from the collaboration? Do the benefits outweigh the risks?
- **Communication:** What are the processes for communicating? Is communication open and constructive? Are there dissenters, and can they be converted to supporters?
- **Commitment:** What is the level of organizational commitment to the partnership and how is this demonstrated? Are partner expectations commensurate with execution capabilities?
- **Continual learning:** What has each partner learned about how to work effectively with another organization?

I have found these questions to be very helpful as I have discussed new collaborative opportunities for Fedora Commons. An example of a very successful corporate partnership that was built upon dialog that addressed these questions is the Fedora Commons partnership with Sun Microsystems. This partnership is very synergistic around shared interests in providing open source solutions to enable preservation and archiving of digital content. It is complementary in that Fedora provides repository

software and Sun provides storage solutions that can underpin the repository. The partnership exhibits excellent interpersonal relationships and communication. Each side provides value in terms of software development, community/customer outreach, and product testing.

We are currently engaging with another non-profit organization, the DSpace Foundation, to evaluate our collaborative potential. Both DSpace and Fedora Commons share a similar mission and are interested in collaboration to help serve our communities better. Both organizations are motivated to show how our open source repositories offer a unique value proposition compared to proprietary solutions. I look forward to reporting to you in the next issue of *HatCheck* on the results of these exciting discussions with DSpace, and other organizations we are working with. Please contact us if you would like to discuss any of the forms of collaboration I have discussed.

***SPECIAL NOTE: DSpace Fedora Discussion
 Tuesday, June 10, 2008***

DSpace Foundation and Fedora Commons will be meeting with a few key folks from each community to begin the discussions on identifying ways the communities can work together to advance the missions of both organizations and communities. This will be the first in many discussions with our communities in the hopes of coming up with tangible ways to easily work together. In our effort to make these discussions as transparent and participatory we have set up a Skypecast. Skypecast is a broadcast of the meeting available for anyone with a Skype account to listen in and participate. The Skypecast can accommodate 100 people. In addition to the Skypecast, we will post the notes from the meeting on our wikis, and hope to have a recording of the meeting that anyone can download.

We realize the technology, language and timing to get massive transparent participation is very limited. However, we promise to continue to make all discussions open and have multiple opportunities to voice your opinion, this being the first of many possible opportunities.

If you would like to listen in (Tuesday, June 10, 2008, 1pm EDT, GMT-5 hours), the meeting will go for 3 or 4 hours and you can log in at:

https://skypecasts.skype.com/skypecasts/skypecast/detailed.html?id_talk=4668237

or search for DSpace at:

<https://skypecasts.skype.com/skypecasts/home>

You must have a Skype account to join the meeting and currently skypecast does not support MAC.

Engineering Punchlist

Fedora 3.0 Release Schedule and Features

By Dan Davis, Chief Software Architect, Fedora Commons

Ithaca, NY — The long-anticipated Fedora 3.0 Beta 2 will be released in early-June, and will be essentially the same as the final public version of Fedora 3.0, which should follow in early July. Fedora 3.0 Beta 2 contains all the features that will be found in the final release and can be used for development testing.

It has been a long road which could not have been accomplished without significant help from the Fedora community that include many commits on the Content Model Architecture plus code contributions. In particular, we would like to note Gert Pedersen and Mogens Sandfaer for updating GSearch, MediaShelf's REST API, Sun Microsystems' ST5800 adapter, and the Fiz Karlsruhe work on scalability testing.

We also delivered a maintenance release, Fedora 2.2.2, in March. Our plans are to continue supporting Fedora 2 for at least a year since there are many production repositories depending on it.

As the startup of the Fedora Commons organization progresses, we are working to become a more community-driven open source organization. We have established licensing standards for code contributions and the first draft of the Fedora Commons Technology Roadmap has been published. The roadmap is a living document which will change with your inputs. As part of our growth, we have organized development into a set of five individually managed projects.

Fedora Repository Project

This project is the traditional Fedora distribution which includes the Fedora Repository service and several closely related components and libraries.

Topaz Project

Based loosely on the ORM (Object to Relational Mapping) family of software, Topaz is a powerful object to triple persistence and query service.

Akubra Project

The Akubra Project is a new effort developed jointly by the Fedora and Topaz project teams as the first area of joint architecture work. The goal of Akubra is to provide a pluggable storage component that supports transactions on common file systems plus the ability to support multiple customized storage options at the same time.

Mulgara Project

Mulgara is one of the premier RDF databases (semantic triplestores) available in open-source and a key component of the Fedora Repository service and Topaz OTM.

Fedora Middleware Project

The goal of this work is to provide an improved Fedora Repository service and to enable new service integrations in accordance

with the principles of Service Oriented Architecture (SOA). This work also includes the integration of “solution bundles” which is an easier path to immediate use.

News

Muradora 1.3 Release

Chi Nguyen

Sydney, AU — Faceted search, all-in-one installation, RSS/Atom feeds, and a thumbnail preview and gallery view are all part of the 1.3 release of Muradora announced by the DRAMA team at Macquarie University in Sydney Australia. Muradora 1.3 is available for download at: <http://www.muradora.org/software>.

Muradora is a web-based graphical user interface (GUI) for the popular Fedora repository platform, built using enterprise Java Spring and Struts 2 frameworks. Among the common features found in a typical repository such as search, browse, self-submission, and versioning supports, Muradora enables flexible access control for end users (based on the XACML standard), inter-domain authentication and federated identity (using Shibboleth implementation of the SAML standard), and multiple metadata schema management (via W3C XForms standard).

More information about Muradora, including deployment and user guides, can be found at: <http://www.muradora.org>.

Notable features in 1.3 release:

Faceted Search

By incorporating GSearch 2.0 with Solr support, users can perform faceted searches, i.e. one can now narrow down search results based on other categories.

All-in-One Installation

There is now an installation script for Unix/Linux systems which will install all the necessary components for Muradora. The complete package is called “muradora-allinone”.

RSS/Atom Feeds

Users can subscribe to collections (even non-public collections) and get notifications of new objects added to those collections.

Thumbnail Preview and Gallery View

Thumbnails are now generated automatically for images. Thanks to the work by the MediaShelf team, one can browse and search using either the traditional listing view or with the gallery view.

Feedback from users is appreciated. Please contact Chi Nguyen (chi@melcoe.mq.edu.au) with queries and suggestions.

Stargazing: Relating Information and Applications to the Data Workflow

Sayeed Choudhury

Baltimore, MD — Whether you are a professional or amateur astronomer, stargazing is a time-honored way to reflect on human endeavor in relation to the vastness of space. The National Virtual Observatory (NVO) is a community-based effort that has resulted in an interoperable framework of data and services, giving people all over the world easy access to data from many different projects at all wavelengths of the electromagnetic spectrum from radio to gamma rays. If an astronomer in Europe analyzes a portion of the sky, another astronomer can compare that analysis with other data sets that depict the same portion of the sky.

The Digital Research and Curation Center (DRCC) at the Sheridan Libraries of Johns Hopkins University is building Fedora-based infrastructure that will curate astronomical data associated with publications. Currently, the high-level, refined data that is generated by analyzing data releases from projects such as the Sloan Digital Sky Survey (SDSS) are distributed over an ad hoc network of personal websites, hard drives or servers. Even worse, many of these high-level data are simply being lost. By building the infrastructure using Fedora, it will be possible to integrate the data acquisition and curation process into existing publishing workflows. Scientists will be able to submit data at the same time that they submit their papers to publishers. Fedora makes it possible to relate information and application to the data workflow.

While the NVO work represents the initial set of development activities, the DRCC is now in negotiations to archive the entire set of SDSS data. SDSS, part of the NVO, began over a decade ago with the visionary goal of providing detailed optical images covering more than a quarter of the night sky, and a 3-dimensional map of about a million galaxies and quasars. Sayeed Choudhury, Associate Dean of University Libraries and Hodson Director of the DRCC reflected on the lessons learned from SDSS and NVO. The entire SDSS data and associated metadata comprises about 100 terabytes. With this scale and complexity, new infrastructure will be necessary as scientists look to libraries for data curation support. Winston Tabb, Sheridan Dean of University Libraries at Johns Hopkins recently remarked, “Data centers are the new library stacks.” Choudhury added, “Without data curation infrastructure, we are losing the fundamental primary research materials that support scholarship.”

Bringing Seven Centuries into the Future: Bibliotheque Nationale de France

Thomas Ledoux

Paris, France — The Bibliotheque Nationale de France had its beginnings in the royal library of Charles V in 1368. The National Library of France (BnF) became the first free public library in

1793 as political change and revolutionary spirit led to an expansion of public access to knowledge.

As many national libraries work towards preserving, maintaining and expanding access to scholarly works and valuable original documents, the BnF has launched an ambitious initiative to expand its digital services by establishing a Distributed Archiving and Preservation System (SPAR) based on the Fedora platform. A pdf version of BnF's use cases and analysis entitled, "Distributed Archiving and Preservation System Bibliotheque Nationale de France" may be downloaded here:
http://bibnum.bnf.fr/spar/CCTP_SPAR-REA_AnnexeI-en.pdf.

The Bibliotheque Nationale de France has chosen Atos Origin to build SPAR. The first release is planned to be delivered at the beginning of 2009.

BnF and Atos Origin are currently defining interfaces for the SPAR system, particularly for ingest features in order to handle packages as opposed to fedora objects (meaning one OAIS package can generate several fedora objects).

For more information please contact Maxime Aimetti or Charlotte Fabre at Atos Origin (<http://www.atosorigin.com/en-us/>), and Thomas Ledoux (thomas.ledoux@bnf.fr) or Jerome Dupont (jerome.dupont@bnf.fr) at BnF.

Spoken Word Services Delivers Digital Audio and Visual Content

Iain Wallace

Glasgow, United Kingdom — Spoken Word Services (SWS)¹ has evolved from a project in the multi-million dollar JISC/NSF funded "Digital Libraries in the Classroom" Programme². The project is a collaboration between Glasgow Caledonian University³ and the BBC⁴ in the UK, and Michigan State⁵ and Northwestern Universities⁶ in the US. Based in the Saltire Centre⁷, at the heart of Glasgow Caledonian University, Spoken Word Services aims to transform and enhance learning by providing access to innovative web based tools, and to a repository containing a wide selection of digital audio and visual content.

In order to enhance access to, and preservation of, the material obtained from the BBC and other sources, SWS has for several years been using repository software – Repos⁸ – developed by project partners at Michigan State University. Requirements for greater flexibility, scalability and semantic awareness led to choosing Fedora for further development. SWS are now in the process of migrating all repository content into a Fedora-based architecture.

SWS are founder members of the Fedora UK & Ireland Users Group⁹ and were hosts for the second group meeting in Glasgow that was held in October 2006. They are also partners with the University of Hull in the JISC funded REMAP project¹⁰, investigat-

ing records management and digital preservation workflow using Fedora repositories and WS-BPEL.

Alongside current 'back end' Fedora developments, SWS is also undertaking a revamp of the 'front end' of its repository based largely on Muradora¹¹, developed by Macquarrie University in Australia¹², and now being used by Hull in the UK as well. Muradora sits alongside Fedora and supports federated identity (via Shibboleth authentication) and flexible authorization (using XACML).

References

1. Spoken Word Services (<http://www.spokenword.ac.uk>)
2. Digital Libraries in the Classroom Programme(http://www.jisc.ac.uk/whatwedo/programmes/programme_dlitc.aspx)
3. Glasgow Caledonian University (<http://www.gcal.ac.uk>)
4. BBC (<http://www.bbc.co.uk>)
5. MATRIX, Michigan State University (<http://matrix.msu.edu/>)
6. Research & Academic Technologies, Northwestern University (<http://www.it.northwestern.edu/about/departments/at/>)
7. The Saltire Centre (<http://www.gcal.ac.uk/thesaltirecentre/>)
8. Repos (<http://matrix.msu.edu/research/repos.php>)
9. Fedora UK & Ireland Users Group (http://www.fedora.info/wiki/index.php/Fedora_UK_and_Ireland_User_Group)
10. Remap (<http://www.hull.ac.uk/remap/>)
11. Muradora (<http://drama.ramp.org.au/cgi-bin/trac.cgi>)
12. Macquarrie University (<http://www.mq.edu.au/>)

Fedora Performance and Scalability Wiki Online

Matthias Razum

Karlsruhe, Germany — It is of paramount importance for large-scale applications that the Fedora Commons repository platform can handle huge amounts of data efficiently. While Fedora is generally known to be stable and reliable, there appears to be a lack of data and experience regarding large-scale installations, and what the related performance implications are. The Fedora Performance and Scalability Wiki Online (<http://fedora.fiz-karlsruhe.de/docs/>) has been established by eSciDoc (<http://www.escidoc.org/>) to investigate and track efforts around these issues.

There are some known performance issues, namely regarding the ingest of new objects. It remains to be investigated if there can be some room for improvement. A faster ingest would be beneficial for many large projects using Fedora Commons. The aim of the Fedora Performance and Scalability Wiki Online therefore is to gather data, document limits and constraints, devise recommendations and help improve Fedora Commons evolve further by contributing the findings to the community.

About eSciDoc

Generating, processing, disseminating, and archiving scientific and scholarly knowledge is changing in a fundamental way. In the future, researchers from all over the world will cooperate, collaborate and communicate throughout the whole process of knowl-

edge generation by means of e-research environments. Typical scenarios include storing, manipulating, enriching, disseminating, and publishing not only the final results of the research process, but all intermediate steps as well, such as pre-research documents, primary and experimental data, pre-prints, and learning materials.

eSciDoc is a joint project of the Max Planck Society and FIZ Karlsruhe, funded by the Federal Ministry of Education and Research (BMBF), aiming to build such an e-research platform for multi-disciplinary research organizations. This not only includes software, but establishing organizational measures, policies, and best practices as well.

Plug a Wiki into a Fedora Repository and Get... A Scholarly Publication

Carol Minton Morris

Ithaca, NY — A new National Science Digital Library (NSDL) scholarly publication, *Classic Articles in Context (CAC)*, was launched in April 2008 with an atmospheric science theme: “Climate Change and Anthropogenic Greenhouse Warming: A Selection of Key Articles, 1824-1995, with Interpretive Essays.” (<http://wiki.nsdli.org/index.php/>) *Classic Articles in Context* will present additional significant scientific questions of the Twentieth Century using landmark and important legacy papers in future issues.

Semantic models for scholarly communications should allow for the creation of new context about published works while mapping relationships to original sources and allowing for materials to be widely discovered and utilized. *Classic Articles in Context (CAC)* does just that by leveraging NSDL's NCore wiki plug-in to its Fedora-based data repository to capitalize on NSDL's ongoing relationships with publishers. *Classic Articles in Context (CAC)* provides publishers with a way to contribute to the creation of new knowledge around published articles in support of teaching and learning.

To illustrate science as a process that builds, and often turns, on discovery and replication expressed in the archival literature of empirical findings, NSDL works with publishers to make the original, full-texts of select “classic” articles available to students whether or not their institution holds a subscription to the journals in which they appeared. Every article featured in a CAC topical concentration includes a narrative essay that provides an overview of the investigation suitable for introductory undergraduate science courses. The essays identify and/or explain particularly significant aspects of the studies (novel methods, for example) and place them within the context of the overall literature of their field (noting, for example, how a given set of findings influenced subsequent work).

The collection currently features 21 papers in full text. Four of the articles, the largest representation among 10 contribut-

ing publishers, appeared in AMS journals during the latter half of the 20th century. Narrative sections include hyperlinks to referenced resources within the National Science Digital Library Data Repository, and in Pathways discipline-specific collections as well as hyperlinks to information about the elements of scholarly publishing.

Currently CAC is being cited by Climate Change Education.Org (<http://www.climatechangeeducation.org/>), and the American Meteorological Society ([http://www.ametsoc.org/NSDL Annotation](http://www.ametsoc.org/NSDLAnnotation)).

Object Reuse and Exchange (OAI-ORE) Standards Beta Release

Carl Lagoze

Ithaca, NY — Over the past eighteen months, the Open Archives Initiative (OAI) (<http://www.openarchives.org/>), in a project called Object Reuse and Exchange (OAI-ORE) (<http://www.openarchives.org/ore/>), has gathered international experts from the publishing, web, library, and eScience community to develop standards for the identification and description of aggregations of online information resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. Although a motivating use case for the work is the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, the intent of the effort is to develop standards that generalize across all web-based information including the increasing popular social networks of “web 2.0”.

The beta version of the OAI-ORE specifications and implementation documents (<http://www.openarchives.org/ore/0.9/toc>) were released to the public on June 2, 2008. These documents describe a data model to introduce aggregations as resources with URIs on the web. They also detail the machine-readable descriptions of aggregations expressed in the popular Atom syndication format, in RDF/XML, and RDFa.

The table of contents page with links to the following other documents is located at <http://www.openarchives.org/ore/toc>.

The full press release for this beta release is located at <http://www.openarchives.org/ore/documents/oreBetaPressRelease.pdf>.

Road Report from Sun PASIG

Carol Minton Morris

San Francisco, CA — According to an IBM study all 1,407,724,920

people who use the Internet¹ will double the amount of digital information² in the world every 11 hours by 2010. Should we keep it all? And if the answer is yes, how will that work exactly? Should it be accessible or will it land in a “dark” archive to be mined by future digital archeologists for its meaning and context?

Sun Microsystems hosted a group of leaders from institutions and organizations worldwide in San Francisco May 23-25, 2008 to consider these and other questions at the Preservation and Archiving Special Interest Group (PASIG)³ meeting. Presentations⁴ were focused on technical and organizational strategies for how burgeoning collections of global knowledge that now include very large data sets might be managed and sustained. Institutional capacity for creating, managing and studying many petabytes of scientific and humanities data has increased as mass storage and information technologies have advanced.

Gail Truman, Business Development Manager, ST5800, and Eric Reid, ST5800 Developer, Sun Microsystems, kicked off the PASIG meeting⁵ with early adopter updates along with opportunities for organizations interested in the ST5800 “Honeycomb” system to take part in an open forum about content addressable storage trends.

A variety of Sun products address both structured and unstructured storage. The Honeycomb creates a storage cloud that gives you a “claim check” to go find data with object identifiers. The Honeycomb box is a “cell” – two cells fit into a rack. The system is designed to allow for services and applications that would normally be accessed in the compute layer to be available within the “cell.” Accomplishing on-the-fly data transformations is one example of what this bundled hardware/software solution allows.

Presentations from PASIG are available here: http://events-at-sun.com/pasig_spring/presentations/. A San Francisco Chronicle article entitled, “Librarians Discuss How to Store World’s Data” by Tom Abate about the meeting can be accessed here⁶.

Humans have never before been challenged with having to save so much information. Historically people have been compelled by circumstance to “let stuff go,” albeit often unwillingly. The list of what many of us leave behind can include almost everything we care about—books, hard drives, memorabilia and artworks—to even bigger items such as houses and cars. Whether selling, donating, recycling, sharing or being forced to abandon our stuff, we are more or less wired to cope with the lifecycle of “things” as they intersect with our lives.

The ancient Alexandria Library, for example, was founded in the third century B.C. and painstakingly developed over time by scribes who hand-copied texts that often arrived by ship. When the library burned scholars and citizens mourned the passing of an irreplaceable collection of knowledge. In 2003 a new Library of Alexandria⁷ was founded based on the same ancient scholarly traditions. The open and unique practice and style of early Egyptian scholarship persists in the new library even if the highly combustible scrolls and artifacts do not. Determining “how”

information is developed and made available, and what policies surround its use may be as significant as “what” the content of the information is in determining the “value” of which data should be preserved for the future.

§ § §

Fedora Commons, the DSpace Federation, EPrints and many other organizations who make use of these and other types of repository software presented overviews of priorities, community developments and technical achievements. The Fedora Commons, DSpace, and EPrints presentations from Sun PASIG are available here⁸.

§ § §

Sandy Payette, Executive Director of Fedora Commons sketched recent Fedora Commons history and reviewed what’s on the short term horizon. Scholarly communications, semantic knowledge spaces, data curation and preservation and archiving are all areas where Fedora Commons has established new and ongoing partnerships to help create technology to sustain the world’s knowledge repositories and archives.

Fedora Commons is establishing community-led Solution Councils for open access publication, data curation, eResearch, and preservation and archiving. In the near term these groups will articulate a mission and develop requirements that will lead to establishing bundles of software and applications to address the needs of organizations whose efforts are focused in these four areas.

What is Fedora right now? Payette reviewed Fedora’s key features which include a digital object model that can:

- Aggregate content “datastreams” in an object... any type of content
- Intermix both local content and external content
- Relationships among digital objects (via RDF)
- Register “content models” for known object patterns

As well as Repository Services:

- Modular
- Web service interfaces (REST/SOAP)
- Versioning
- Dynamic service binding based on object content model types
- File-centric (all essential characteristics in XML files)
- RDF-based indexing (semantic triplestore/index with query)
- Security with pluggable authentication and XACML policies
- Journaling (replay all events to create replicas of repository)

§ § §

Andrew Treolar presented an overview of the Australian National Data Service (ANDS)⁹. This is a multi-institution, long-term and very large data pilot project established to ensure that research-

ers are able to identify, locate, access and analyze any available research data. They are looking at issues such as multi-discipline access and use outside of research. Australian funding agencies are tracking the implications of the ANDS work so that they can potentially require a data management plan as a part of new research grants.

§ § §

Michelle Kimpton, executive director, DSpace Federation¹⁰ gave a report on the recent release of DSpace 1.5 which uses the Maven build system. This distributed development tool supports “publishing models.” This functionality is useful for those who are looking to use DSpace “beyond institutional repositories.”

The DSpace Federation has changed the model for how they work with their community in order to manage growth. Almost anyone can now contribute code right away because builders do not contribute code back to the core.

Manakin is part of DSpace 1.5, and is the DSpace default user interface. Manakin allows for customized UIs that look more like a digital library. A number of organizations are using Manakin as a front end to their repositories. More information is available in this December 2007 article from DLib Magazine: “Manakin, A New Face for DSpace”¹¹.

§ § §

Dave Tarrant, PRESERV¹² and the University of Southampton, introduced a storage controller for EPrints repository software¹³. The controller supports a pluggable storage layer for repositories, providing the ability to store objects in different locations based on metadata or type, and enables direct interaction between the repository software and open storage platforms. Tarrant’s presentation, “From Open Storage to Smart Storage: Enabling EPrints Repository Preservation” can be accessed here¹⁴.

References

1. <http://www.internetworldstats.com/stats.htm>
2. <http://techbloom.blogspot.com/2007/02/total-amount-of-digital-information-on.html>
3. <http://www.sun-pasig.org/>
4. http://events-at-sun.com/pasig_spring/presentations/
5. https://events-at-sun.com/pasig_spring/agenda.html
6. <http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2008/05/29/BU3710U908.DTL>
7. <http://www.bibalex.org/English/index.aspx>
8. https://events-at-sun.com/pasig_spring/presentations/
9. <http://www.pfc.org.au/bin/view/Main/Data>
10. <http://www.dspace.org/>
11. <http://www.dlib.org/dlib/november07/phillips/11phillips.html>
12. <http://www.preserv.org.uk/>
13. <http://www.eprints.org/>
14. <http://eprints.ecs.soton.ac.uk/15818/>

Upcoming Events eSciDoc Days 2008 in Berlin

Karlsruhe, Germany — The Max Planck Digital Library and FIZ Karlsruhe cordially invite you to participate in eSciDoc Days 2008 (<http://www.escidoc.org/JSPWiki/en/EscidocDays>), to be held Monday, June 9 and Tuesday, June 10, 2008 at Harnack Haus (<http://www.harnackhaus-berlin.mpg.de/eng-index.htm>), Berlin, Germany which is traditional venue for presentation and discussion of German academic research.

The eSciDoc Days are targeted at both eSciDoc users and communities interested in digital libraries, repository systems, e-research infrastructures, and (scientific) knowledge management.

During eSciDoc Days extensive information will be provided about the aims, results and challenges of the project, both from a functional and a technical view. Presentations and tutorials by eSciDoc team members as well as by eSciDoc users (from MPG and external institutions) will demonstrate the current eSciDoc Solutions and the underlying eSciDoc Infrastructure, which is based on the Fedora Commons open source repository platform.

We have reserved time for in-depth discussion of questions brought up by the participants. Please see the preliminary programme at <http://www.escidoc.org/JSPWiki/en/EscidocDaysProgramme> for more information. Registration information is available at <http://www.escidoc.org/JSPWiki/en/EscidocDaysRegistration>.

Schedule of Summertime Fedora Commons Community Meetings and Events

Ithaca, NY — Fedora users are planning meetings and events in Canada and Europe throughout the summer and early fall. Take a look at upcoming opportunities to share ideas and information. Contact organizers for information about how to participate.

- Dutch Fedora Users will meet on June 19. Contact Lodewijk Bogaards <Lodewijk.Bogaards@dans.knaw.nl>.
- Scandanavian Fedora Users will meet on June 24-25. Contact Eva Müller <Eva.Muller@kb.se>.
- UK & Ireland Fedora Users will meet on July 8. Contact Richard Green <R.Green@hull.ac.uk> or Chris Awre <c.awre@hull.ac.uk>.
- The Red Island Repository Institute will be held on Prince Edward Island August 11-15, 2008. This 1-week hands-on workshop will be led by Sandy Payette, Fedora Commons Executive Director; Richard Green, Manager, RIDIR, REMAP and RepoM-Man Projects, e-SIG, Academic Services, University of Hull, and; Matt Zumwalt, MediaShelf. Contact Mark Leggott <ri@upei>.

ca>.

- A Fedora EU inaugural event is planned for Sept. 19 in coordination with the 12th European Conference on Digital Libraries (ECDL) to be held in Aarhus, Denmark. Event details will be available soon.

Contact David Flanders <d.flanders@bbk.ac.uk>.

Of Repository Road Shows and Fringe Festivals

The JISC Common Repository Interfaces Group (CRIG) is sponsoring a Roadshow due to head out on July 12 for a 2 1/2 week, multi-city event called “CRIG Roadshow.” The event will feature developers who will be pitching and sharing on-the-spot ideas in hopes of convincing whoever is on hand to participate in expressing CRIG’s collaborative slogan: “The coolest thing to do with your data will be thought of by someone else.”

A repository “fringe” event will be coordinated with the well-known Edinburgh Festival Fringe to be held August 3-25, 2008 during the “Preview Week” of the Edinburgh Fringe Festival. Many of the festival’s two thousand comedy, theatre, music, opera and dance shows will be starting their month-long runs with tickets at half price. All repository researchers, managers, administrators and developers are invited to come and join in Edinburgh’s newest and entirely unofficial festival!

Carnegie Foundation Open Education Webevent

You are invited to join a webevent at the Carnegie Foundation Teaching (<http://commons.carnegiefoundation.org/>) and Learning Commons focusing on open education resources, open source development, and open knowledge to be held June 16-19, 2008. More information is available at <http://commons.carnegiefoundation.org/webevent>.

Advocates of open education resources, open source development, and open knowledge building all hope—even promise—to use and build on information in the public domain. They do so in order to equalize access to information and educational opportunities for learners all over the world. Sponsors and individuals also build collaborative communities to create knowledge, objects and information for the public good, using information technology to spread their tools and resources widely. With such high expectations, we must ask:

- Is it working? How do we know? When will we know?
- Is the “open education movement” filled with overwhelming optimism?
- Is the movement changing the way we learn and how knowledge and resources are created, or is open education another form of patronage, whereby knowledge and information travel in one direction?

Please join this interactive webevent to discuss these issues and share your hopes and fears about open education. We invite you and your colleagues to explore the Teaching & Learning Commons’ resources, and share your own links to related work. Help build understanding of open education by exchanging your projects, thoughts and critiques in an open, online discussion. Join the conversation beginning June 16.

Bookmarks

Download the “Roadmap” to Fedora Development

Ithaca, NY — The Fedora Commons core development team has made a Roadmap to development available at <http://www.fedora-commons.org/resources/roadmap.php>.

The Roadmap summarizes the strategic vision of Fedora Commons that guides development, documents community themes and priorities that summarize the needs we will address in the document, and provides a plan for future software releases to aid in planning for those who adopt the Fedora platform.

This is a draft and a living document. In the future it will be moved into a new Fedora Commons Wiki as an interactive service to keep users, developers, vendors, and decision makers informed about progress towards software releases, and provide community access to development tools that will encourage participation. But don’t wait, please join the process and send your comments to the fedora-commons-users mailing list on SourceForge.net. Sign up for the list at <https://lists.sourceforge.net/lists/listinfo/fedora-commons-users>.

NSDL Director Kaye Howe at JA-SIG, “Ubiquity, Interdependence, and the Age of Collaboration”

Saint Paul, Minnesota was the site of JA-SIG 2008 April 27-30, 2008. With the theme, “Higher Education Solutions: The Community Source Way” it’s no surprise that National Science Digital Library (NSDL) Core Integration Director Kaye Howe was on hand to deliver the closing keynote address entitled, “Ubiquity, Interdependence, and the Age of Collaboration.” She noted that Prometheus discovered both the power and danger of knowledge when he stole fire from the gods for use by mere mortals. “There is no going back once you have stolen fire,” she said.

“Education is the great activity of the fallen world,” she said. “It is what we need to create a just society,” and yet, she pointed out, “We think of all the difficulties and all the ways we have to reinvent this wheel over and over again.”

Howe is no stranger to pointing out that the legacy of working together to create widely accessible opportunities for education has been one of conflict. These initiatives have at their core

the conundrum of both striving for perfection while building on imperfect systems. She noted that the power of collaboration has at its roots a rich scholarly history and tradition with well-known relationships such as the public and private lives of John Adams and Thomas Jefferson as examples of a stormy collaboration with a remarkably successful outcome.

She concluded by noting that “The imperfect is our paradise,” as Wallace Stevens suggests in “The Poems of Our Climate.”

THE POEMS OF OUR CLIMATE

Wallace Stevens

I
Clear water in a brilliant bowl,
Pink and white carnations. The light
In the room more like a snowy air,
Reflecting snow. A newly-fallen snow
At the end of winter when afternoons return.
Pink and white carnations - one desires
So much more than that. The day itself
Is simplified: a bowl of white,
Cold, a cold porcelain, low and round,
With nothing more than the carnations there.

II
Say even that this complete simplicity
Stripped one of all one's torments, concealed
The evilly compounded, vital I
And made it fresh in a world of white,
A world of clear water, brilliant-edged,
Still one would want more, one would need more,
More than a world of white and snowy scents.

III
There would still remain the never-resting mind,
So that one would want to escape, come back
To what had been so long composed.
The imperfect is our paradise.
Note that, in this bitterness, delight,
Since the imperfect is so hot in us,
Lies in flawed words and stubborn sounds.

Fedora Commons Quick Poll I Results: User Gatherings

Ithaca, NY — There were 21 responses to five short questions on this quick poll that asked the Fedora community for OR08 impressions, and also about face-to-face gatherings in general. The poll went out on April 18, 2008.

A few nuggets:

I. 57.9% of respondents have attended other local Fedora User Group Meetings.

2. “Preservation and Management” and “Architecture” were tied in the “Extremely useful” OR08 session topics category with 57.1% each.

3. The winner for “In the future I would like to attend” is “A multi-day, international Focus on Fedora conference” in Europe—66.7%.

4. Fully 80% of respondents would take the time to attend either a “Using Fedora” web seminar or an “Informational Fedora web seminar.” Discussion and perhaps some additional testing?

5. 58.8% of respondents thought that Fedora Day at OR08 content was excellent.

Look for more Fedora Commons quick polls on a variety of topics. Contact Carol Minton Morris (clt6@cornell.edu) with questions or suggestions.

Strands of Global Web Knowledge Come Together at Third International Open Repositories Conference 2008

It was cool and sunny earlier this spring as over 450 delegates from 35 countries made their way to the University of Southampton in the UK where the School of Electronics and Computer Science was set to host the Third International Open Repositories Conference. This year's conference, held April 1-4, 2008, was sponsored by U.S. and European companies and organizations including Sun Microsystems, Electronic and Computer Science, University of Southampton, JISC, Texas Digital Library, Web Science Research Initiative, Repositories Support Project, Jorum, The Depot, and Enovation Solutions. Conference organizers from three open source software repository groups – EPrints¹, Dspace², and Fedora Commons³ – were on hand to present solutions, and to exchange ideas with attendees in day-long user group meetings and during informal poster sessions and receptions.

The theme “Practice and Innovation” was evident in the main conference program as well as in meetings and events held in conjunction with OR08, as presenters told stories of experimentation often coupled with the challenges of running day-to-day repository services for their companies and institutions. In sessions focused on interoperability, legal, models, architecture and frameworks, national and international perspectives, scientific repositories, social networking, sustainability issues, usage, and Web 2.0, presenters shared both groundbreaking and tried and true use cases, solutions, and results of inquiries for how repositories can better meet the needs of the global scholarly and scientific communities that they serve.

The conference opened with an observation by conference co-chair Les Carr, University of Southampton. He suggested that the collective efforts of the delegates – repository managers, librarians, archivists, developers, project leaders and representatives

from IT companies – who are working directly with worldwide research and information producers are creating a global web of knowledge.

Carr and his University of Southampton team have made all OR08 conference proceedings available in, not surprisingly, a repository. You may browse and download presentations and posters here: <<http://pubs.or08.ecs.soton.ac.uk/view/subjects>>.

Peter Murray-Rust, a Reader in Molecular Informatics at the University of Cambridge and Senior Research Fellow of Churchill College, gave the keynote address. He began by asking how many of the assembled attendees worked in laboratories. A small number of hands went up. “We have a data drought (in laboratories),” He said. “Permission barriers caused 90% by people and only 10% by technology are preventing direct access to up-to-date scientific research findings.” Murray-Rust suggested that most discovery takes place along the long tail of science, among small groups of laboratory scientists who need access to current data. He also noted that there is a vast difference between the technical service needs of bench scientists versus those of theoretical scientists. “You can’t sit in a building with ivy growing up it and be removed from scientists in laboratories with their test tubes and small furry animals and call yourself a ‘scientific repository,’” he said. Murray-Rust gave several examples of technologies and services that would help scientists in laboratories. His suggestions included RSS systems that email “active molecules” harvested and updated hourly by robots; incentives that compel scientists to contribute; text mining technologies that reclaim “lost” data from PDFs; getting involved in scientific data collection workflow “upstream” closer to where data is created; and developing access to petabyte stores at universities so that there are facilities to hold scientific data at the source.

Session Highlights

Dean Krafft, Cornell University, gave a presentation in the Interoperability session on NCore, an open-source platform for creating digital libraries united by a common data model and interoperable applications. Built on top of Fedora, NCore includes a suite of library management tools and services, as well as a variety of end-user tools for the collaborative creation of context and content. The NCore system already supports the National Science Digital Library (NSDL), implementing one of the largest production Fedora repositories in existence, and it is being considered for other new digital library implementations. Krafft’s presentation was significant in scope and impact, and provided attendees with multiple ideas for extending off-the-shelf Web 2.0 tools to integrate with metadata and structured content in a digital library framework.

Sustainability affects what happens over time to virtually all aspects of personnel and technology associated with ongoing repository operations. Stuart Haber began the Sustainability sessions at OR08 with a presentation entitled, “A Content Integrity Service for Digital Repositories.” He suggests that it is critical to be able to verify that a document, or piece of content, is what it

claims to be. His system creates a “witness” for each piece of a document – paragraph, sentence, speech, for example. The “witness” then computes a certificate of authenticity that is coupled and stored with the original document.

Mary Marilino, director of e-Science and the NCAR Library and former director of DLESE, spoke frankly from personal experience about what to do when the email arrives informing you that your funding will not continue. She explained her planning for sustainability and asked that people think openly about what happens when and if a funding agency might “pull the plug.” She introduced DLESE, The Digital Library for Earth System Education – one of the first big digital library projects in the U.S. DLESE was a grassroots, community-led project with a completed collection of 13,500 digital educational resources organized into 41 thematic collections. As an organization DLESE was a focal point for community action in geoscience education and developed numerous best practices around building education-based digital libraries.

The digital artifact is now being sustained in partnership with several groups – NSDL, DLS, and NCAR. DLESE will remain a valuable educational resource, and an important partner in NSDL. Although the community governance element of DLESE is no longer funded, Marilino hopes that it will remain very much a vital community focal point, and anticipates continued active dialog with the community of DLESE users.

Les Carr presented “End of Life Scenarios for the Repositories of Virtual Organizations.” In giving this talk he hoped that he would not be known as the “man who burns repositories.” The alternative title was “Or: who cleans up when the party ends?” His investigation focused on whether a repository could be sufficiently simplified into a low-cost static website once its community of users disappear along with their requirements for active services. Collecting and curating over time is what a persistent and permanent repository backed by policies and institutional commitment implies – this activity is not intended to be a fly-by-night dumping ground, even though events, organizations and funding streams are always in flux.

How old is old? How persistent is persistent? Carr talked about venerable institutions like the University of Oxford, for example, which existed in 1096 and perhaps even earlier. Seats of learning are by their nature institutions that can be counted on to last. Repositories created by virtual organizations that come about as temporary collaborations between independent partners, on the other hand, may last about a decade or less. Virtual organizations often come into being as a part of grant activities without the benefit of clear institutional affiliation. Carr suggests that the “squillions of dollars” spent on international, highly collaborative projects may ultimately equate to longer repository lifespans. If the institution that backs the repository, however, disappears, contents are often tied up in administrative and resource allocation knots, leaving information consumers without access unless inexpensive solutions are available.

Around the Edges

Conference organizers worked with the JISC CRIG (Common Repository Interface Group) to design and support “The Repository Challenge.” Nineteen small teams of developers competed during OR08 to achieve goals set by repository managers and the user community. The teams were hard at work into the night on each day of the conference. The OR08 web site states, “We think this (Repository Challenge) is the first time ever that an effort has been made to get rapid cross-platform international development happening in the same place and at the same time with some of the best repository developers in the world!”

Winners were Tim Brody, University of Southampton; and team members Ben O’Steen, University of Oxford and Dave Tarrant, University of Southampton, who modeled layered Fedora Commons and Dspace t-shirts while accepting the award at Thursday evening’s banquet. Their application was ‘Mining the ORE’ (Object Reuse and Exchange Protocol). More information about the Repository Challenge may be found in the JISC Information Environment Team blog.⁴ Interactive flip-charts from OR08’s Repository Challenge teams that illustrate highly collaborative and creative engineering processes may be found in CRIG’s Flickr photostream.⁵

Open Archives Object Reuse and Exchange (ORE) at OR08

The OAI-ORE UK Open Day⁶ funded by JISC was the main event during day four. Conference delegates were there to find out about the latest improvements to the third alpha release of ORE “specifications that allow distributed repositories to exchange information about their constituent digital objects.”

Herbert Van de Sompel, Research Library, Los Alamos National Laboratory opened the meeting with background and motivations for creating a new resource-centric infrastructure. He was followed by Carl Lagoze, Computing and Information Science, Cornell University, who described the underlying data model and gave an overview of implementation in Atom. Michael Nelson, Computer Science, Old Dominion University, talked about discovery of resource maps. Community members Robert Sanderson, University of Liverpool, Jim Downing, University of Cambridge Center for Molecular Informatics, and Thomas Place, DARE and University of Tilburg, gave quick overviews of their experiments with ORE. During the afternoon session Simeon Warner, Computing and Information Science, Cornell University, focused on practical issues including proxies, relationships, lineage, nesting, and implementation in RDF/XML and RDFa.

Web artifacts are seldom single entities. An individual web page, for example, might be comprised of multiple bits of text, images, and media in several formats that together form a unit of related information on the internet. While information managers are in the business of developing applications that collect sets of related information like web pages, ORE standards provide a way to describe and exchange whole “constellations” of complex aggregations of scattered and varied information using a widely

applicable description of sets of objects.

Save the Date

The Fourth International Open Repositories Conference 2009 will be hosted by the Georgia Institute of Technology in Atlanta May 18-21, 2009. Check for more information as it becomes available at <<http://openrepositories.org/>>.

References

1. Eprints, <<http://www.eprints.org/>>.
2. DSpace, <<http://www.dspace.org/>>.
3. Fedora Commons, <<http://fedora-commons.org/>>.
4. JISC Information Environment Team Blog, <<http://infteam.jiscinvolve.org/2008/04/18/open-repositories-2008-2/>>.
5. CRIG’s Flickr photostream, <<http://www.flickr.com/photos/wocrig/>>.
6. OAI-ORE UK Open Day, <<http://www.openarchives.org/ore/meetings/Soton/agenda.htm>>.

Who is Using Fedora?

(continued from page 1)

Being able to provide an expanding picture of community vitality to the world demonstrates to potential users and funders that Fedora Commons is providing sustainable technologies to create, manage, publish, share and preserve digital content as a basis for intellectual, organizational, scientific and cultural heritage worldwide. It is also important to help users of Fedora find each other in order to leverage initiatives and avoid duplication of efforts.

The next question is, “What are they doing with it?” The Registry is designed so that each entry is a link to a wiki page where each institution can provide more information about what they are doing with Fedora.

Currently the Registry is a wiki page. The next iteration will be database-driven. We are working on an access solution to make it easy for users to give themselves wiki accounts again, without leaving the door open for spammers. In the meantime a wiki account for a project, organization or institution contact person will be assigned when you register. Additional accounts may be requested here.

Please send feedback about the design and function of the registry, or content updates to Thornton Staples (tstaples@fedora-commons.org).