



FedoraCommons

Fedora international users' meeting

Open Repositories 2008

Southampton, UK

3rd & 4th April 2008

Thursday 3rd April, 0900-1030

Plenary session

Introduction and welcome

Sandy Payette, Fedora Commons

Keeping alert: issues to know today for long-term digital preservation with repositories

Neil Beagrie, Charles Beagrie Ltd

Neil is a leading expert on digital preservation and curation and has an international reputation across the archive, library, science and research sectors in the long-term management of digital assets. He has played a major role in developing strategies for digital preservation including founding the Digital Preservation Coalition and overseeing development of the UK Digital Curation Centre: he was also the international consultant to the US National Digital Information Infrastructure and Preservation Program (NDIIPP). His career to date has spanned a range of senior information management roles at the British Library, the JISC, the Arts and Humanities Data Service, and the Royal Commission on the Historical Monuments of England.

Neil will reflect current digital preservation activities and trends to inform what repository developers and managers need to know today to ensure effective long-term digital preservation.

Fedora Commons: a technical update

Sandy Payette and Eddie Shin, Fedora Commons

Case studies

Putting the ORA into FEDORA

Sally Rumsey

This paper will describe from a repository manager's perspective how Oxford University Library Services (OULS) is using Fedora to build an institutional repository, Oxford University Research Archive (ORA), for research materials. It will explain the reasons for selecting Fedora and how it addresses the need to provide a system that meets the requirements of the many diverse groups that comprise this large and complex research-intensive institution. Fedora has been selected as the basis for a bigger architecture of repository services: ORA forms only one element of this repository landscape.

Implementing an Integrated DAMS - Fedora/VITAL at the National Library of Wales

Paul Bevan and Glen Robson

The National Library of Wales (NLW) is currently undertaking a National Digital Asset Management programme centred around Fedora. Building upon the VITAL solution, the NLW has committed to a programme of digital object management through a single system architecture which will provide the support for the library's own digital activity over time.

The NLW has an extremely diverse collection reflecting societal and technological changes across Welsh history. The collections include a range of born-digital objects as well as the products of a flourishing digitisation programme.

This paper outlines the developments which the National Library has undertaken in order to ensure that the digital and digitised collections of Wales are securely managed in perpetuity. Through a discussion of the issues and processes involved, this paper will draw out key themes from the NLW experience and provide insights for other Fedora sites on the different ways in which the DAMS can be implemented and represented.

Repository redux: UPEI VRE

Mark Leggott

The Robertson Library at the University of Prince Edward Island recently embarked on a comprehensive development program for the stewardship of information resources in the three primary academic "landscapes": administration, learning and research. This paper presents our experience developing a collaborative research environment using the open source Drupal and Fedora systems as the two primary components, including technical details of how the two systems were integrated. A comparison of our "Repository Redux" approach with more traditional models for institutional repositories, which have not always met with the anticipated success, will be discussed. The session will also touch on aspects of capacity building and staff development at a small academic institution which are enabled by the use of open source applications.

Preservation and management

SOAPI - a flexible toolkit for implementing ingest and preservation workflows

Mark Hedges

The digital material being held in repositories is increasing significantly not only in size but also in diversity of format and complexity of structure. This is particularly the case for repositories based on Fedora, with its highly flexible content model architecture. This data represents a major investment, and in many cases is irreplaceable; there is thus a pressing need for implementing preservation functionality in repositories to ensure long-term access, and for scalability reasons this must be as automated as possible.

The SOAPI project is addressing this need by developing a flexible toolkit for implementing automated and semi-automated workflows supporting the preservation of complex digital material in repositories. Our approach is not to develop a monolithic tool, but a set of modular web services, each encapsulating a well-defined unit of preservation functionality, which can be configured and combined to produce workflows.

Records management and digital preservation using Fedora and iCalendar

Richard Green, Chris Awre and Iain Wallace

The eSIG team at the University of Hull, working with partners Spoken Word Services at Glasgow Caledonian University, has been fortunate to secure funding from the JISC for the REMAP Project. REMAP takes up more or less where Hull's successful RepoMMan Project left off and exploits the BPEL-orchestrated Web Services environment to add value to digital objects at the point they are moved from a user's private space to the outward-facing area of the institutional repository. In particular, the REMAP toolset will build into this process a system for ongoing records management and digital preservation (RMDP). The system will use the established iCalendar standard to flag the need for future management and preservation events.

Introducing a batch metadata editing tool for digital repositories: FABULOUS

Jenny Quilliam and Prashant Pandey

In 2006 the ARROW (Australian Research Repositories Online to the World) Project funded two mini-projects at the University of South Australia to solve the issue of batch metadata editing. The combined outcome of this work resulted in the creation of web based, open source PHP application called FABULOUS (Fedora ARROW Batch Utility with Lots of User Services). FABULOUS reduces time required to manage repository content by enabling modifications to user selected sets of objects and data streams.

The presentation will cover the background and the solution to an important issue faced by repository managers on a day to day basis.

Architecture**Fedora and Honeycomb: A new buzz in creating and managing large-scale digital archives***Eric Reid*

Rapid growth of digital content is driving the need for a new approach to preserving and managing archived data. With vast stores of data that must be preserved for many decades to come, many organizations are finding it costly to maintain the integrity of their data and are challenged by the complexity of storing content in a way that enables users to easily find and retrieve what they need. Sun has partnered with Fedora Commons to create a solution that can help organizations store, manipulate, and access their digital resources in a flexible and cost-effective manner. This solution combines the Fedora open source software platform with the Sun StorageTek™ 5800 system to provide a petabyte-scale object store that greatly simplifies the task of preserving massive amounts of data over long periods of time.

This paper will describe the direction Sun and Fedora Commons are taking with this promising combination of technologies. We will review the work done to date, the expected long-term advantages to this approach, current Proof of Concept deployments, as well as future plans.

(repository +/- e-Infrastructure) ?!*Andreas Aschenbrenner, Tobias Blanke, Mark Hedges and Frank Schwichtenberg*

Various initiatives are moving towards shared, virtualised hardware resources for their repositories. - Could these (isolated) approaches converge just like federation activities are converging? What can e-Science concepts and grid technologies, which have virtualisation at their very basis, contribute to this?

As the somewhat cryptic title indicates, this paper aims to trigger discussion. The Fedora community has already experimented in this regard, and - technologically - Fedora is arguably best suitable to venture further into this. -- We identify steps towards Fedora gridification. However, apart from technology, what about user trust?, and who takes responsibility for those requirements, that may exceed current capabilities of (national) e-Infrastructure?

The architecture of the Oxford University Research Archive*Benjamin O'Steen*

This paper will describe the architecture and construction of a digital archive, powered by Open Source software such as Fedora-Commons, Apache Solr and Darwin Calendar server. The architecture that will be described is the same as the one underpinning the Oxford University Research Archive (ORA) and, in future, other repositories such as the Forced Migration Online archive. It will detail the choices made, with emphasis on the utility of certain chosen conventions, along with some of the technical details of the implementation.

Datasets**Video conference capture, cataloguing and storage in a web services framework***Alistair Young*

UHI Millennium Institute provides university level education within the Highlands and Islands of Scotland, covering an area roughly the size of Belgium and therefore relies rather heavily on video conferencing to deliver teaching and learning across the thirteen widely distributed partner campuses.

Many lectures are delivered using video conferencing and it was felt by the teaching staff that archiving select video conferences would be an ideal way to make them available online as teaching resources, either within the institutional virtual learning environment (VLE) or as resources in their own right. Adding standard metadata, such as Dublin Core, would also allow tagging and searching of these new learning objects by UHI staff and students respectively. To that end, this paper describes the confluence of two projects, VCR and CTREP: the goal is to develop an end to end capture, cataloguing and viewing system, with a backend Codian providing the recorded VCs, Fedora providing the metadata and access control and an industrial strength streaming server, affectionately known as The Toaster, providing long term storage, with Fedora mediating access to that storage via a user interface provided by Sakai.

NJVID – A New Jersey statewide video portal based on Fedora*Ron Jantz, Grace Agnew and Isaiah Beard*

NJVID is an Institute of Museum and Library Services (IMLS) grant funded project that will provide a statewide digital video portal to meet the wide ranging needs of educators, students, and lifelong learners in New Jersey. A digital video portal for the state of New Jersey must support the complex educational needs of K20 schools as well as the diversity of information and cultural heritage institutions of which there are over 600 including public libraries, archives, historical societies, and museums.

Video collections to serve the citizens of New Jersey are the core of NJVID. Collections will focus in three areas: Video Commons will be publicly available including history, lectures from notables, and video documenting research and scientific advances. Commercial videos will be available through educational consortia and accessible based on the rights of participating institutions. In a special application, lectures on demand will be supported by enabling faculty to create virtual clips that can be used for the duration of a course.

The last paper in this group was withdrawn which gave us the chance to fit in a late submission from Fez albeit not on datasets.

Cool mashups and new features with the latest Fez+Fedora - media streaming, Solr powered search, Facebook integration and more

Christiaan Kortekaas

Fez powered Fedora repositories are now more flexible than ever. Content can be imported, deduplicated, exported, deleted, undeleted, tombstoned, embedded, streamed, fed, discovered and managed in some innovative new ways.

The Fez developer community has contributed some very exciting modules recently including Solr integration, commenting and rating of objects, Origami flash-based image viewing (e.g. for high resolution images), and a 'back in time' view of datastreams based on Fedora versioning. These together with new streaming 'youtube-like' video and audio disseminations, custom themes per community, Facebook integration, Endnote importing, and smart 'fuzzy logic' object deduplication features will be detailed and demonstrated during the presentation.

Of interest to FedoraGSearch2 users will be our Fez implementation of Solr, which is the first of its type where authorisation filtering is performed during the search, thereby solving the notorious 'paging' problem hampering other Lucene post-query filtering approaches. Utilising Solr's advanced faceting, full-text highlighting and other Lucene-based features brings highly scalable and searching to content.

The presentation concludes with a look at the future of Fez and Fedora development at UQ and the growing contribution from the Fez developer community.

Search

An SRW/U-compliant search service for Fedora

Michael Hoppe and Matthias Razum

Distributed object search and discovery across multiple repositories requires some interoperability standard. Search & Retrieve for Web/URL Service (SRW/U) is a powerful and flexible example for such a standard. We have implemented a SRW/U-compliant indexing and search service, which is part of the eSciDoc Infrastructure. It works with arbitrary metadata profiles and supports indexing of several file formats, including PDF and Microsoft Word. Even though the service is part of eSciDoc, it can be configured to work directly with Fedora. We will describe its architecture and demonstrate its adaptability. The software is available as open-source software.

Fedora in the world of integrated search

Asger Blekinge-Rasmussen and Kåre Fiedler Christiansen

For most of the users of Fedora for managing digital objects, Fedora is just one of the sources of digital information available to them. Thus you will often see a library or archive having several available databases for searching, each with their own user interface and query language. Some places will offer a federated search option: You can write one query string, it will be submitted to multiple databases, and the search results of each will be presented. In contrast, integrated search is the art of searching multiple, heterogeneous data sources using just one index. User studies at The State and University Library show that most library users never look at anything but the main search service at the library. Thus if the digital material in the Fedora repository is not available through the main search service, most of the users will never see it.

This presentation will show how Fedora can without great effort integrate with the open source integrated search engine Summa, using an atomistic data model, the proai OAI-PMH provider, disseminators, web services and XSLT technology.

An investigation into filtering of search results by access constraints

Gert Schmeltz Pedersen and Christian Tønberg

This work is investigating the problem of filtering of search results in order to filter out hits that the user does not have access to see according to the authentication and authorization constraints of the Fedora repository that was queried.

The investigation was triggered by a discussion on the fedora-commons-developers mail list. The RAMP project (<http://drama.ramp.org.au/>) brings Fedora GSearch into the same security context as Fedora, so that XACML policy enforcement can be done during query evaluation. The Fez project (<http://sourceforge.net/projects/fez/>) uses a combination of a simplified XACML engine and a relational database.

This investigation aims to clarify the options and their properties and, through the application of a set of use cases it throws light over, how developers and repository managers could and would select among the options, depending on their requirements.

Programming

Fedora + Atom

Edwin Shin

This paper proposes extending Fedora to support the Atom Syndication Format in two different contexts: as a digital object serialization format and as a messaging format. Fedora currently supports two serialization formats for digital objects: FOXML and METS. The translation of the Fedora Object Model to Atom is possible with little or no extension to Atom, where an Atom feed represents a Fedora object and each Atom entry represents a Fedora datastream or datastream version. By supporting Atom as a serialization format, the variety of tools and libraries available for processing Atom suddenly become tools for processing Fedora objects. As a deposit format, Atom promises to improve Fedora's interoperability with other, non-Fedora repositories by providing a repository-neutral format with widespread tool support across multiple platforms.

Fedora 3.0 introduces a messaging service to provide notification of repository events. The messaging idiom of topics and queues neatly translates to the Atom idiom of feeds, with each event represented as an Atom entry. The widespread adoption of Atom has also resulted in an abundance of Atom clients. Using Atom as a messaging format enables, for example, the repository manager to use a web browser or email client to track repository events.

Fedora & XForms: Current uses, future potential

Kevin Ford

Based on work at Columbia College Chicago developing Cloche, the Library's custom front-end application for its Fedora repository, this paper will illuminate general XForm development issues specific to Fedora digital object XML formats. Additionally, it will identify current and promising avenues of use for XForms with Fedora, both for administrative purposes (XACML potential?, e.g.) and for cataloguing work.

How we integrated Fedora into Ruby On Rails, and how you can use it

Matt Zumwalt

Breathe a sigh of relief, and let yourself daydream a bit. Developing Fedora-centric applications just became a lot easier.

MediaShelf (<http://yourmediashelf.com>) has now created software libraries that allow developers to treat Fedora repository content as Ruby objects. Even better, we have made it so you can use these objects natively within Ruby on Rails. Matt Zumwalt, MediaShelf's Lead Architect, will demonstrate how this graceful union was achieved.

Matt will start by reviewing Ruby on Rails fundamentals and showing how the new RubyFedora and ActiveRepository libraries fit into that framework. He will then walk you through a simple Rails application that uses the libraries. The session will conclude with a look at some of the underlying code and a discussion of the current development status for this work.

Accommodating diverse search requirements over a Fedora repository

Michael Durbin and Jon Dunn

Fedora offers the flexibility to store all types of digital objects with any type of metadata. In the Fedora repository for Indiana University's Digital Library Program, intended to serve both access and preservation needs, we store an increasing number of object types representing items ranging from scanned books to documentary photographs to three-dimensional puzzles. Other object types, including digital video and audio, are also on the horizon. In addition, we store objects representing more abstract concepts such as a piece of sheet music for which various editions may have been scanned, or an archival collection containing many component objects. Metadata useful for finding a given item may exist at any of these levels, and many items also contain searchable full text content. Besides simply having a wide variety of objects, collection purposes, access models, and user needs vary widely across the contents of the repository. With such diversity, the common ground between objects shrinks, as does the usefulness of a one-size-fits-all search solution.

In this talk, we will cover strategies for managing such diverse and evolving discovery needs over a growing Fedora repository. We will discuss our experiences with existing tools as well as the design and justification for new tools that we have developed. With sustainability as one of our primary goals, we will discuss the technical and organizational challenges associated with the incremental process of providing useful and complete search and discovery access to our Fedora repository from within a variety of user environments.

Friday 4th April, 0900-1030

Front ends

Easy on Fedora - Experiences with eSciDoc *Lodewijk Bogaards*

Easy On Fedora is a project of DANS that uses relatively new and unique software called eSciDoc, a logic tier on top of Fedora. In a short presentation Lodewijk Bogaards, project leader of Easy On Fedora, will present to you our up-to-date experiences with eSciDoc.

Plone as a front-end to the Fedora repository *Peter Reimer*

The Fedora repository software is a core component of the Open-Access publication platform developed by the Digital Peer Publishing Initiative (DiPP). Starting with 8 there are now 14 publication projects from various disciplines using the well established format of an electronic journal. The platform is run and developed by the Library Service Centre (hbz) of NRW.

Fedora was chosen as a service layer due to its scalability, extensibility and the availability of an OAI interface. Since Fedora lacks a user friendly GUI, the Content Management System Plone, based on the web application server Zope was selected to build the front end of the publication system. Special content types have been developed, which store its content in the Fedora repository instead of Zope's own database. Plone/Zope itself is written in Python.

In the presentation we will show how the Python modules were generated, how the functionality of the repository is made available in the content management system and what other services are connected via web services, like URN assignment, document conversion, etc. Finally, a few examples of Open Access Journal using our platform are presented.

Muradora: A turnkey Fedora GUI supporting heterogeneous metadata, federated identity, and flexible access control

Chi Nguyen and James Dalziel

In this paper, we present Muradora, a new turnkey GUI for Fedora that addresses the need for federated identity and flexible access control. Muradora incorporates a new authorization architecture that moves the access control enforcement back to Fedora. This allows users to access the back-end repository via multiple interfaces, but still having the same consistent access control across all of them. Muradora also presents a new standard-based framework for metadata validation and input that allows new metadata supports without modification of the core GUI code.

Semantic technologies

Building personal collections and networks of digital objects in a Fedora repository using VUE

Anoop Kumar and Nikolai Schwertner

Digital repository technology has evolved over the past decades enabling streamlined creation and sharing of digital content. Current implementations make it feasible to setup repository infrastructures not only for large institutions but also for smaller groups and individual users. Yet, one of the difficulties in the adoption of the technology remains the relative scarcity of powerful tools for publishing content to digital repositories. With this in mind, we developed a Fedora publishing module allowing instant ingestion of VUE-generated content maps in a Fedora repository. We take advantage of the relationship facilities within Fedora to build a knowledge base of interrelated VUE maps and resources though user contributions.

Content-based image retrieval integrated into Fedora

Pierre-Yves Burgi and Patrick Monbaron

We designed a content-based image retrieval system built on the MPEG-7 standard and the Fedora architecture. Initially, the bank of images with their associated metadata was handled through *ad hoc* application software developed at the University of Geneva. The idea was thus to migrate the whole bank of images onto the Fedora system while affecting the least possible user interfaces.

The second part of the project was about building a user interface for image retrieval. Search process is typically initiated with users performing textual searches on metadata (based on gSearch), which eventually result into lists of images matching textual data. Then, users have the possibility to search for additional images based this time on content similarity. Content similarity involves distance measurements performed on the MPEG-7 descriptors represented as data vectors built from the textual data stored in Lucene.

Using Fedora in the European project TAS3 to handle person-related information

Marc Santos and Ingo Dahn

Person-related information has to be handled in a large variety of contexts. The European Integrated Project "Trusted Architecture for Securely Shared Services (TAS3)" aims to support trusted handling of such information with a system of specially designed web services, using Fedora instances as its standard repositories.

In order to deal with the variety of data format specifications in use in the sectors of education, employability and health, TAS3 follows a specification neutral approach, i.e. data are stored in repositories in an internal format. Only for delivery they will be transformed on the fly to satisfy the requirements of the specification in use by the receiving party. This will be realized by accessing data objects through format-specific Fedora disseminators.

Within Fedora, complex data will be distributed to simple interrelated data objects, following the Core Components approach of UN/Cefact. This will facilitate the integration of legacy databases into the framework developed in the project. Specific Fedora data streams will be used as proxies, keeping aggregated XML representations of complex data objects.