# DSpace Fedora Collaboration

Contents

## Introduction

On June 10, 2008, DSpace and Fedora representatives met to discuss possible collaborations.

From Fedora: Sandy Payette, Dan Davis, Chris Wilper, Thorny Staples

From DSpace: Michele Kimpton, Richard Rogers, John Erickson, Robert Tansley, Bradley McLean

The meeting was skypecast, and most of the meeting was recorded.

## Action items/outcomes

1. DSpace and Fedora teams will define pilot projects to prove out the usefulness and capabilities of ORE, with possible integration with Zotero as the end user application.
2. Sandy/Michele to evaluate the CRIG roadshow as an opportunity to do advocacy on the collaboration, prototype coding and communications. The is week of July 28th in DC.
3. Michele/Sandy will provide analysis of DSpace and Fedora roadmaps, noting timing, overlaps, and potential synergies
4. Michele/Sandy write up a report of the list of collaborative ideas (all the great ideas that came out of the meeting) - broadcast to communities and figure out way to get feedback
5. Michele/Sandy talk to organizations running both Fedora and DSpace to find out most compelling projects; and maybe some others evaluating both?

## List of possible ideas generated during the meeting

1. Identify high impact scholarly applications (e.g., Zotero) to integrate with our repositories; initiate meetings to move forward
2. Jointly mentor the Google+Summer+of+Code project that will demonstrate DSpace running on top of Fedora
3. Define common content models;
   a. Map the DSpace data model to Fedora digital object content models;
   b. Define a common reference data model for institutional repositories
4. Investigate adopting common standards (protocols; interfaces; formats) for deposit of content into our repositories (e.g., SWORD, APP, other)
5. Investigate scenarios for integrating a open source workflow engines with our repositories; demonstrate workflows that use both DSpace and Fedora
6. Define and implement a common event notification architecture (e.g., JMS provider in open source)
7. Define and implement common storage APIs (e.g., JSR170, XAM, Fedora's Akubra) for interfacing repositories to underlying storage systems
8. Joint implementation of the ORE data model to enable sharing and exchange objects from our repositories
9. Work out issues around the semantics for content models using ORE (e.g., journal article, image, journal, book, etc.)
10. Develop shared services/modules to enable exchange of objects among repositories using ORE?
11. Host a web service for creating personal aggregations/collections using ORE?
12. Joint support of JHOVE; investigate possible funding strategies to provide support to JHOVE initiative
13. Investigate use cases from both communities to come up with shared output ??
14. Investigate shared user interface approaches
    a. Manakin on top of Fedora and DSpace?
    b. What is best way for to provide an out-of-box repository application that works with both DSpace and Fedora
    c. Other ways to enable building of lightweight apps on top of our repositories?
15. Move toward common architecture for modularization of core software; evolve to a plug-in approach where modules are sharable by both DSpace and Fedora?

## Notes: DSpace+Fedora+Collaboration

### Michelle Introduction

Morning: Familiarization on both platforms

Richard Rodgers presented Make DSpace more modular. Inflow of code and patches not well supported.

New information model which can handle improved structure

Flexible workflow

Maven being used to build now.

Based on recommendations from over a year ago.

Decentralized development but it is hard to take on more complex, larger tasks without centralized help. Example is major change to information model and supporting core code.

Committer driven process but is a lot of work including documentation and testing.

When no organization started brought together a core team to coordinate some of the major modifications using community to do coding. To complete in Feb-Mar timeframe.

## Sandy Introduction

Formed Fedora Commons about a year ago to provide a home for Fedora and collaborations

Initial Roadmap published and uses an Eclipse like approach. From key contributors and project architects and conferences outputs.

Must get repositories behind what is actually going on and must support where the Web at large is going. And improved storage approaches.

Themes: Interface with the Web much better (e.g. new REST API; Sword API). Ways to provide easier update as Solution Bundles. Collaborate with DSpace, ATOM, OAI-ORE are examples of how to provide better defined service interfaces. Better plug-in architecture for core Repository Service (e.g., Spring)

Integration with distributed services Web, Web Services.

Workflows, service orchestrations. Difficult but want to find low complexity approach which is good enough outside the Repository.

Introduction of Content Models to aid in usability/creation of objects.

5 Employee Committers, addional 10 Committers, 10-15 Contributers.

Thorny Staples forming Solution Councils to create Solution Bundles.

## Brainstorm on Collaboration

Seed Question: Why are people excited about collaboration?

Thorny - Conceptual semantics. Getting our assumptions together. Explicitly state assumptions. Shared idea about content models and use cases. Collective wisdom.

Sandy - Look forward to where the world is going. Anticipate the future directions that form our place in the new world. What is our role?

Richard - Mission Statement.

Sandy, Thorny - More of a conceptual framework. Sustain the use of data.

Sandy - Not low hanging fruit more ongoing process.

Brad - Practical on the ground services needed now. But support where world needs more than one model.

Thorny - For now, just a first cut conceptualization.

Sandy - Can we do some common services which both communities can use. Especially for exchange.

Brad - Service and protocols implementation could serve as a model for customizations.

Peter Murray - Consider focusing on Web (e.g. REST) API instead of Web Service (SOAP) implementations. Maybe not Sword. [I'm pretty sure I didn't say this. I'm definitely sure I didn't mean to say it. Peter Murray 23:09, 22 June 2008 (EDT) ]

John - Protocols too low, need to be problem focused. Solve needs for real users who cannot accomplish what they want to do. Look for users with both Fedora and DSpace. Steward Lewis. Uncover problems and solutions.

Thorny - Workflows may be a good place to understand needs. Hull worked with BPEL and has added People extensions.

John - Imperial College integration with Enterprise IS is example.

Sandy - Submission workflow is a common example.

Dan - Also very light web, Web orchestration solutions for distributed organization often virtual organizations.

Sandy - Big ingest.

Brad - Workflows to move between repositories, into repositories.

Thorny - We need to the services to do the steps and we share them in common. This is an obvious low hanging fruit.

Richard - Exposing repository to Workflow not embed.

Thorny - But it needs to be coherent and have good integrations.

Michelle - What about Manakin on Fedora?

Thorny - It would fit the solution bundle approach.

Sandy - Synopsis of Google+Summer+of+Code

Rich - It is ambitious, but can be scoped to be doable in 8 week timeframe. Basic goal is provide a DSpace interface on top of Fedora backend.

Brad - look really hard at the ingest process; where are commonalities

John - New DAO+layer has been added to DSpace which may provide an interface layer.

Rich - Push data and metadata pushed to Fedora instead may be more practical.

Sandy - Think about collections too to show the graph.

John - The work should be done incrementally.

General thinking it is good to store relationships particularly collections.

Rich - High value for consciousness is DSpace over Fedora (or vice versa). Look at both roadmaps. Workflow looks good. OAI-ORE. Storage. SWORD. Atom expression of Sword deposit.

John - Policy. Binding policies to objects, workflows, services. New to users but needed. No model on how to do management in an efficient way. Need stakeholder community to drive the needs for this. Will be hard. Can start now but is a long term project.

Sandy - Look at Mulgara, McQuarie University, Chi Nguyen. In middleware, outside repository.

John - Federation. Loosely couple autonomous repositories. Virtual subject collection.

Sandy - Demand has not been well described. Except Search.

Rich - Replication networks.

John - Transformations.

Peter Murray (Skype) - Dark Archives

Dan - Can be addressed incrementally, positioning versus Cloud.

Rich, Chris - Storage to pass more than just blobs.

John - What about XAM.

Sandy - When will XAM be done.

John - There is a debate who will do XAM, vendors or open source. Vendors want not do dumb down storage.

John - we can aim for a repository storage model that maps well to XAM storage model.

Dan - What about support for Hierarchical storage.

Rob - Common event notification. Enterprise event queue.

Michelle - JSR 170?

Sandy - Levels get increasingly harded.

Chris - Xpath part is hard.

John - JBoss DNA new definition of Federation based on supergraph model.

Chris - Sharing of content models. Community registry of content models.

Sandy - How to introspect on ORE.

Dan - Common requirements for components, services, suggestions to funders, registry of availability, download support location.

John - Semantic interoperability.

## Ways to Move Forward on Collaboration

Sandy - Digest the ideas. What are the next steps.

Michelle - Must be transparent and communicated broadly.

Rob - I missed his statement.

Brad - Google SOC. Ingest process.

Sandy - A roadmap of the collaboration.

John - would like to see an overlay of timelines in terms of roadmaps; analyze roadmaps; come up with a roadmap for the collaboration. Look at near term plans and see what seem to shake to top; also put the options out to the community.

Sandy - We can offer a confluence/Jira space for DSpace and Fedora Collaboration provided by Fedora Commons infrastructure

Action Item - Create a DSpace Fedora Wiki Space. Get everyone logins.

John - how would Forester or Gartner characterize what's going on here. We would have to characterize the market, the stakeholders, the trend, the features. Who's plaiying – the open repositories, but also Jackrabbit, alfresco, cloud, ad hoc solutions. Who can do this? HP is trying to do it for repository federation space. It's hard to do, especially how players address the future.

Michelle - Create a list and prioritize or order possible collaborations.

John - Characterize the various repositories and features to describe the market both open source, cloud, Jackrabbit, commercial and map to current /future needs. Plus why the need items (e.g. specific protocols & formats). Find out want our users actually want.

Michelle - Must have a lightweight agile way to satisfy needs.

Brad - View this is a layered perspective, not just software.

Michele - we should focus on the open source mission in general, beyond just the technology. Keep focused on the mission. We by collaboration we can have more people focus on the mission and less on the tactics.

John - if the mission is properly tuned, we should be able to execute it without being beholder to our existing codebases.

Sandy - We should watch out for the "innovators dilemma" where successful, well managed companies were trumped by disruptive technology (citation to Clayton Christiansen's book).

John - Make sure we are clear about the mission and ensure it looks far enough into the future. Innovator's dilemma makes this point; can execute mission well and land your plane in the mountain. Thorny made the point - what's it really about? Is it the long term preservation of bits? The long term access to information. We need to understand what that mission needs to be.

Thorny, Sandy - Get a clear idea who we are serving.

Michele - reminder about the Repository Roadshow (Ben O'Steen and Dave Tarrant gig that's touring the USA soon). Sandy and Michele can be present at such events ("shaking hands on the steps of the capital") 🙂 . We should identify venue to bring the DSpace and Fedora communities together (e.g. barcamps, BOFs, etc.

Action - Find more joint venues for Sandy, Michelle, and DSpace/Fedora communities.

Closing action item - Michele, Sandy, Brad, and Thorny will have a beer together to summarize ideas generated in this meeting and conceive of next steps.

# Joint projects

A strategy session was held in August 2008 to come up with a list of short term and longer term projects. Further definition on each of these can be found at DSpace Fedora jointprojects