

# Fedora 3 to 4 migration hold ups

Hydra Connect 2015 Thursday Morning Unconference Session

Migrating from Fedora 3 to 4

Informal Poll: about half a dozen people plan to migrate in the next six months

Started with questions and filled in as many answers as we could.

Pain points? What is holding you back?

- What to do with the FC3 Datastreams – convert to RDF or keep your XML as a proxy?  
ANSWER: Moving to RDF will be a big win, so worth doing. If you have terabytes of XML, you have a migration problem and an underlying metadata problem. Experience: ContentDM to RDF presented as a cataloging problem. Metadata working group is looking at ways to handle this. Flat XML schema will be easier than a more complex schema (MODS). RDF will be faster. Can also keep a read-only copy of the XML, but don't try to synch them up – historical artifact.
- What effect will adopting FC4 have on performance with RDF instead of XML, esp. for write operations.  
ANSWER: Writing should be similar, reading may be slower. There's less potential for conflict in write operations. There is work going on to optimize performance now.
- FC3 distinctions will disappear – instead of categories (administrative metadata, descriptive metadata, etc.) metadata will be one big bucket – how to manage? How to differentiate rights metadata vs. descriptive metadata, for editing, for metadata dumping.  
ANSWER: can handle this in the application layer – list things you can edit and things you can't – using Presenters for this in latest code. Not really a Fedora concern.
- How to use the projected file systems of FC4 with ActiveFedora – is this the right way to do the equivalent of externally managed datastreams (files) from FC3?
- Legacy data conversion – reindexing takes 3 weeks, so need to get it right the first time?
- How to / whether to preserve/transfer FC3 PIDs to FC4, especially for RESTful interface.  
ANSWERS: 1. Keep the PID as the ID by using NOIDs in FC4, to keep the PID as the ID. 2. You can also migrate the PID not as the FC4 identifier, but map it as a property. (DC:identifier has been tried, there were some issues). Fedora 3 namespace still exists in Fedora 4 but has no property for FC3 PID. Some work to be done to find a property to use for this – need to build consensus. 3. Use a different layer of the stack to map from FC4 ID to FC3 PID (e.g. Nginx mapping)
- Interim state – how to handle transfer period – some apps on FC3 some apps on FC4 – possibly the Fedora 4 migration tool is faster than the Penn State gem for migration. Gem transforms bitstreams to native RDF properties. FCRepoMigrate doesn't scale
- Transactionality – are batch updates possible? Will they be possible in ActiveFedora to help with migration?
- What are the best practices for managing a migration – collection by collection? How to do quality control?
- How to migrate the history/versions when you can't revert to an old FC3 version of an object once that object is in FC4? Artifactual info? Audit history?
- Finding/hiring help is difficult.
- High demand for documenting experiences, roadblocks, solutions  
ANSWER: create a documentation working group
- Could try standing up a Fedora 4 instance for a new project to gain familiarity with it before committing to a migration.
- Storage considerations for migrating managed datastreams, esp. if you have terabytes of existing data. – is it possible without doubling your disk space? That would be the best practice but it may not be possible – possibly chunk the data, possibly project from FC4 over an FC3 filesystem?

ACTION ITEMS:

1. Create Documentation Working Group – start the process within a week - send out appeal to the hydra-tech list, get a page going on the Wiki - this will be done by \* David Chandek-Stark  
Other volunteers for this WG: Jim Coble, Mike ??? Andrew ???

2. Adam Wead's migration tool takes the hydra-rights metadata and translates it into Web ACL objects. If you use the Fedora migrator, you'll need to do a second pass to handle this hydra-specific piece.

Andrew Woods: Adam Wead's migration tool predates the Web ACL final implementation. Need to verify that the tool is consonant with the current FC4 implementation of Web ACLs