2018-12-13 Designing a Migration Path Meeting

Time/Place

Time: 9am - 3pm Eastern Standard Time US (UTC-5)

Place: Building 38A, room B1N30Q, National Library of Medicine

Audio Conference: There is a Polycom speakerphone in the room, with phone number 301.827.7446.

Directions

For directions to NIH, maps and security information, see https://www.nih.gov/about-nih/visitor-information. If driving, you must enter at NIH Gateway Drive which is available from Rockville Pike southbound only. Nearby visitor parking is available for a fee. If taking public transportation, we are at the Medical Center stop on the Red Line.

The conference room is in Building 38A, room B1N30Q. Building 38A (Lister Hill Center) is the tall ten-story building adjacent to Building 38 (National Library of Medicine), a three-story building with a pagoda-like roof. Once inside the main lobby of building 38A, walk downstairs to level B1. At the bottom of the stairs, turn right and walk to the end of the hall into B1N30. Turn right, and conference room B1N30Q is on the right.

Attendees

- Erin Tripp
- David Wilcox
- Andrew Woods
- Este Pope
- Scott Prater
- Mark Jordan
- Sayeed Choudhury (I will need to leave at 11:45)
- Timothy Shearer

Absent

- Andy Weidner
- Michael J. Giarlo

Agenda/ Notes

Designing a Migration Path - collection of project resources

Time	Topic	Lead
9:00 - 10:15	Environmental scan review and feedback	David
10:15 - 10:30	Break	

10:30 - 11:45	Survey Discussions
	
	• Research Questions
	  ◦ e.g. What would it take you to migrate/upgrade to Fedora 4/5?
	  ◦ Get rankings and priorities, e.g. How much do you care about the repository back end of your service? How much to you care about the discovery layer of your service? how much to you care about the UI of your service?
	• Scope and scope creep. What do we really need to know? What can we find out that will be actionable? Do we need demographic info about size of repo or institution or budget? Do we need to categorize users?
	• How do barriers like budget, technology spec, product loyalty weigh in your decision?
	• Who do we want to talk to? Devs? Admins/DMs? Librarians? All Fedora users? Or just Fedora users who are super involved?
	• Cognitive Interviews - plans to set up calls to test survey in early Jan within and outside of advisory group. It will go through survey and discuss your impressions of it, e.g. did it get at what I wanted it to? Is there a question I missed? Is it too long? etc.
	• Dissemination
	  ◦ Personal recommendations for people we can talk to - advisory members as advocates.
	  ◦ Know our sample set - can we get all emails of people signed up to Fedora Google Groups and on our registry? And email them directly? That way we can know how many people it reached and a percentage of responses. Good metric.
		Erin

11:45 - 1:00	Lunch	

1:00 - 2:30	Dissemination Plans • Open Repositories and other proposals and deadlines • Interim Communications - Fedora newsletters.	
2:30 - 3:00	Scheduling preliminary consultations with advisory boards January next steps - tech review	

Notes

Introductions

Tim - Using Fedora for IR and special collections. Have been migrating for 3 years. Trying to do a lot in the migration process. Special collections will be Fedora 5.0 and we'll roll our own. We're use Hyrax for the IR. We're heavily oriented to preservation and less to access.

Este - We're happy with Fedora 3. Limited dev staff. We're waiting for Fedora 4 and CLAW. We've been talking about migration for 2 years. We're going to leapfrog into 5 or whenever modeshape is gone. We're not ready to be the first implementers. I don't see that represented in the environmental scan. We also run a super early version of Hydra. Hosting everything in AWS in future.

Mark - in Islandora the 'many members' problem isn't an issue as it is in Fedora because of how Drupal is interacting.

Erin - **tech adoption lifecycle.** I see less willingness to go first.

Scott - going first. Before the digital services we're core services. Now they are core services and we can't take as many risks.

Tim - more willing to cut a check than dedicate FTE devs to that. We also have a lot more content to move now than we did have 5 years ago. Less need for exploration leads to a culture where going first isn't desireabe.

David - this isn't represented in the literature.

Este - Karen asked me why I wasn't using DSpace or CONTENTdm. We are a small org. But we do have big collections that are valuable and are being viewed internationally. A migration could disrupt service. We can't be flippant about a move. It's becoming more important.

Tim - Faculty have special needs and the formats are all over the place and that doesn't fit into DSpace.

Scott - **The technological side is missing from the environmental scan.** I talk to storage and development people. I hear a lot about complexity and no one wants to be first and performance issues and configuration. List serve traffics is lots of messages about 'I'm one person and I need to stand this up fast.' Another motivation for us is an **old product and unsupported software**.

David - this grant will help us validate the reasons and get the data we need to support other related projects. We want to know what the best return on investment will be. We might have a lot of surprising information in this grant.

Mark - we migrated from CONTENTdm to Islandora 3 years ago. 1.3 million objects. We created the Move to Islandora Kit. Lots of metadata profiles. We will lose people if the migration isn't reliable. The process is driven by Drupal tooling not Fedora tooling. It's a different route. Anxiety in the Islandora community around Fedora 3 to the new models in Fedroa 4 - the way MODS or any XML metadata to RDF. We have an active metadata interest group. The data model in Fedora 3 - people are used to it and comfortable with it. It's ike an email with attachments. People can't see how that will translate to Fedora 4. Echoing the software is easy and people are hard. **Communications piece here. Do people think about the Fedora. What do people know about versions and difference between. Categorizations.Prepare people to migrate every 3-5 years. How will that impact decision making? Framing migration as an ongoing process that's iterative between the 3-5 year process.**

What about service providers? Do we need those? Users have money and not people. We're seeing that trend.

Tim - I could look at our file system and recreate it. There's comfort in that.

Andrew - Islandora tech migration framework is easy with Drupal. The conceptualizing the models and schemas is hard.

Mark - config files for this process is scary. It's xpath. We can predict many of those but because Islandora is configurable and customizable. So it makes this far more complex. We want a standard configuration that will deal with 70% of the

ACTION - Find out when we will apply for the next IMLS pre-proposal.

Scott - Digital library analyst. Tech lead and strategy and PM work. We're migrating to Fedora 3 right now. It's our dirty little secret! We were one of the first who put up digital content in 1990's. They are going to unplug our servers in 2019 so we have to move. No one owned it until now. 3.5 Million objects. But heterogeneity of the content is the problem. We have all formats. Our IR is DSpace. Just hired Atmire to upgrade. Our concerns are digital preservation and I had a lot of concerns with Fedora 4-5 with the file system. We're really interested in programming around OCFL. **Andrew and Scott will folllow up on this ACTION.** We have been using our own custom discovery system that integrates our digital collections and ALMA, etc. Our Fedora is really a back end system. We don't use Fedora as part of discovery. Our data models are custom and are not reusable. Fedora 3 is easy to learn from an IT perspective. Anorther concern I have about Fedora 4-5 is the triple store. It feels like a career on it's own. If 3 million objects turn into 100 million triples and my triple store falls down.

Andrew - we spoke yesterday about the key components of digital preservation.

Tim - as we started looking at going to Fedora 4-5 Ben Pennell did some testing to find ways to reduce the number of triples. There are ways to bring those down but there are trade offs.

ACTION - We can include the assessment in the analysis portion of the project in January.

Scott - we're talking about investing in getting new content. We need this to scale.

Andrew - building toward this survey.

Scott - could we publish an article as part of project so some of the topics that aren't in the literature are present. Or a blog.

Andrew - echoing David I want to understand if this is technical barriers or something else.

Sayeed - I'm a part of the IMLS board and what I'm doing here has nothing to do with that board. This is based on my experiences at Hopkins. We have gone through a series of migrations from DSpace 1.6. We're looking to move to Islandora CLAW. One point that's relevant to this grant is that Fedora 4-5-6 won't be that different. But people outside this room are uncomfortaable with versioning. Where are the user stories and use cases. Can we have an institutional persona that include the user stories and give them a path. We can't pick all of the things to do after this grant but we could pick one or two. Sense of exploration has decreased. I used to be in charge of an R&D unit with no operational responsibilities. We don't have that anymore. We have in our time wrestled with IT and we can still be explorative but the capacity has decreased. **I haven't heard a lot about data in this conversation. It's a place where we can be innovative.** There are huge vendors out there like Clarivate and others. They say, why do it yourself? With data it's still the wild west and the services haven't been perfected.

Scott - research data is a big interest and that's where the money is going.

Environmental Scan

David - want to get a sense of what people are challenges with and why they are migrating.

Started by going through each source, did a summary, and then documented relevant take aways from each source. Then put together a list of common themes. And created the one page summary.

Migration/upgrade projects are common. The literature includes the motivation, pain points (relevant) and the benefits (some unexpected - different from motivation). Literature covers advice and requirements. Requirements could be useful for our purposes. Some notes talk about the current status of Fedora. There were some notes on tools/resources as well. Tools were most commonly related to metadata processing.

Motivations

- Getting away from commercial products. License limits. Costs
- Performance and scale
- Lack of flexibility (file types, metadata formats, etc)
- Digital preservation support and linked data support but less common than the above.

Difficulties

- Metadata quality, clean up, deduplication, supplementation, transformation to other schemas.

Advice

- Planning
- Metadata normalization
- Migration
- Verification

Stakeholder communications, requirements gathering, scope management.

Agile methodologies recommended. Staff turn over and single points of failure. Needed to define roles and create contingency planning.

Pause for discussion.

Survey Questions

- Importance of asking why institutions want to migrate
    - If institutions are uninterested in migrating to Fedora 4/5 we need to know that
    - Are institutions moving from Fedora to something else?
    - Motivations may elicit more responses than barriers
        - Key repository requirements
- What layer of the stack is highest priority? Discovery, UI, etc.
- When was the last upgrade?
- How often are you willing/able to do an upgrade/migration?
- Who makes/influences this decision at your institution?
    - It may be unclear who makes/owns the decisions, which can itself be a barrier
- If you knew the migration would buy you X years before needing to do another migration would that influence your decision?
- Would the migration be worthwhile if it resulted in new skillsets that can be applied elsewhere?
    - What if it resulted in a nicer interface? (Not as relevant with Fedora)
    - What new affordances will the migration deliver?
- How many objects? TBs? Object models?
- What type of institution?
    - What resources do you have to migrate and maintain a system? Developers, sys admins, etc?
    - IT profile, capacity in house (how many resources can you dedicate?)

- We can get institutional data online
- What is your role? How much influence do you have over the decision to migrate?
- If you're currently on Fedora 3, what do you like about it? What do you dislike?
- What do you like/dislike about Fedora 4/5?
- If you're still using Fedora 3, why haven't you migrated yet?
- Ask a geographic question so we know where the respondent's university is and include institution name so we can dedup.

Things to consider

- Need to know how many people we're sending the survey to. Track percentage of responses
- Working with someone to draft and approve appropriate language for international audiences
  - The focus of the grant is US institutions but we still want to reach out to the global community
  - US institutions are impacted by activities and decisions in the global Fedora community
- Has there been any interest in maintaining and building on Fedora 3?
  - Are some institutions happy with Fedora 3 and uninterested in moving?
- Who to include
  - Service providers? Clients will follow their lead
  - Decision makers
  - Committers
- What does a migration cost? Can have a huge impact depending on how often migrations are done
- Survey should be 10-15 minutes max
- Interviews should help identify gaps/issues with survey.

Ideas

- Insider knowledge? communities of practice.
- Are you willing to be the first? Tech adoption cycle.
- Importance of institutional and IT profiles. Categorizations
- Old product and unsupported software.
- Communications piece here.

  - Do people think about the Fedora.
  - What do people know about versions and difference between.
  - Anxiety of data models and schemas.
  - Prepare people to migrate every 3-5 years. How will that impact decision making?
  - Framing migration as an ongoing process that's iterative between the 3-5 year process.
- What about service providers? Do we need those? Users have money and not people. We're seeing that trend.
- Is tooling and training the best value? In terms of metadata and triples. Skill development.
- Size and complexity of the content. TBs and # of objects and content types data models (compounds) variety of metadata. Do you have content /complexity = small, medium, large. Important for diagnosis.

Cognitive Interviews. Volunteers:

- Este
- Mike
- Scott
- Tim (or someone at UNC)

Front End Application reviews. Potential ones:

- Samvara
  - See action items
- Islandora

  - See action items

Data

- Check with Doren

Existing migration tooling

- see grant for others
- Bridge 2 Haiku
- Metro consortium in NYC - Diego Pino?

Dissemination plan:

- OR? Yes let's try.
- CNI Spring meeting?
- DLF fall meeting?
- Preservation venues? PASIG NDSA SAA
- Partner/co-present with dspace folks?
- Publish (OR paper? C4L paper?)

- Channels for communication?
  - Newsletter already there
- Survey
  - be sure to follow up with recipients.
  - dissemination - targeted.
  - personalize the ask where possible, amplify with folks who have followers?

Potential ways to leverage results

- Could be used to create new proposals:
  - Methodology and resourcing to help institutions get to a less precarious repository state
  - Build migration tools and kits
  - your idea here...
- Transparency about what's going on in the repository world
  - archetyping barriers to upgrading repositories
  - Busting myths (if there are any) so that there is clarity about those barriers

Actions

- Michael J. Giarloidentify potential front end candidates from Samvara community. Prefer PCDM and things that lean generalizable,
- Este Pope Poll community to look for Islandora folks who can be candidates. David Keiser-Clark?
- OCFL review Andrew Woods
- API spec review David Wilcox
- ALL: See conference check list and have everyone indicate what