

2019-04-12 Designing a Migration Path Meeting

Time/Place

Time: 9am - 3pm Central Daylight Time US (UTC-5)

Place:

Washington Boulevard Executive Center
4818 Washington Boulevard
St. Louis, MO 63108

Audio Conference:

Join Zoom Meeting
<https://duraspace.zoom.us/j/5769796461>

One tap mobile
+16699006833,,5769796461# US (San Jose)
+16468769923,,5769796461# US (New York)

Dial by your location
+1 669 900 6833 US (San Jose)
+1 646 876 9923 US (New York)
Meeting ID: 576 979 6461
Find your local number: <https://zoom.us/u/afJIZ8XWA>

Directions

TBD

Attendees

- [Erin Tripp](#)
- [David Wilcox](#)
- [Andrew Woods](#)
- [Este Pope](#)
- [Scott Prater](#)

Absent

- [Michael J. Giarlo](#) (calling in)
- [Mark Jordan](#) (calling in)
- [Timothy Shearer](#)
- [Sayeed Choudhury](#)

Agenda/ Notes

[Designing a Migration Path](#) - collection of project resources

Time	Topic	Lead
9:00-9:15	Welcome, agenda review	
9:15-10:15	Discussion. 'Ah Ha!' moments after reading. Any surprises? <ul style="list-style-type: none">• Designing a Migration Path - Institutional Profiles• Designing a Migration Path - Migration Tool Review• Designing a Migration Path - Fedora API Review• Designing a Migration Path - Oxford Common File Layout Review	David
10:15-10:30	Break	

10:30 - 11:45	<p>Survey distribution timeline (one month behind sched) and strategy</p> <p>Idea to deliver a webinar as project update and to kick off survey distribution.</p> <p>Draft survey preamble: https://docs.google.com/document/d/1vsg4NGJ5aDfwQfp63JG4jCgXtyAb37xnfq6K4_BUs5k/edit</p> <p>Distribution to lists: Samvera-Tech, Samvera-Community, Digital-Curation, Code4Lib, DLF, PASIG, Fedora-Community, Islandora-Community, Islandora-Dev.</p> <p>Personal solicitations for survey from advisory group members. Pick 5 ppl to send message to. List here: https://docs.google.com/spreadsheets/d/1cKerrKyryyyoM9SU6Uhdw7ROpreJgwlvSzMm9TQ7GTU/edit?usp=sharing</p> <p>Timeline for keeping survey open: Two weeks? With reminders during period. Schedule and draft text and assignments.</p>	Erin
11:45 - 1:00	Lunch	
1:00 - 2:30	<p>Planning next steps:</p> <p>Secondary consultation with Samvera, Islandora and Fedora Communities after survey results come back. Advisory board members will receive a summary of anonymized survey results. We will speak to each advisory board member if there are gaps in the data set (e.g. things we thought we would see that we didn't) and surprises in the data set. Feedback will be documented and included in the final report.</p> <p>Action- provide timeline for setting up individual meetings with Advisory Board members</p> <p>Action - Determine timeline for next series of virtual meetings in May/June. Get it on our calendars.</p>	Erin
2:30 - 3:00	<p>Final Report / Report out at OR in June 2019</p> <p>Discussion:</p> <ul style="list-style-type: none"> What needs to be in there for it to be useful? For you, what would mean success or failure? How can we use our findings to pursue an IMLS project grant in Aug/September? <ul style="list-style-type: none"> What about UI/UX migrations. Important to collection managers Who will be at OR in Germany? Want to participate in presentation? Any other events where we can propose presentations delivered by advisory board members? <ul style="list-style-type: none"> DLF call for proposals due end of April. Scott, Este, David all going. 	David

Notes

Agenda review

Erin shared that a practicum student is starting on Monday and will be helping with the grant. Erin will send out a message introducing her.

Overview of documents

Broad overview of documents, AHA moments

David - Reached out to a number of institutions and did a brief survey of front-end data model perspective. What is the same, what is different, how difficult to do a migration based on the front-end framework. Islandora is probably the easiest - homogenous, core structure with solution packs, Islandora already has a migration framework. Some issues, some things to be rewritten, but Islandora will likely have the least trouble. Content will be okay, but custom front end will need to be required. Florida State University, National Library of Medicine, University of Wisconsin-Madison, UNC Chapel Hill, Michigan State University, Stanford University, Williams College, Amherst College

Este - UX/custom front ends are not necessarily critical to migration, but could be major areas of need for many institutions because that is what users and non-tech staff think of as the library interface.

Erin - is this something we need to call out as a priority for migration?

Scott - ORM - object relational mapping may be an important component to consider to allow for more easily developing an interface for the underlying repositories. Could be challenging to build out UI template because of so many data models. But an ORM could be a middle layer to help with editing.

Erin - brainstorm what the next steps are and include this issue of UX layer in the planning.

Tim - the barriers are not all technology barriers. Can be custom metadata, data models. We should at least write down the barriers. Could be to tell some folks that they are a DSpace shop, maybe best to move off of Fedora.

Erin - need to acknowledge these barriers exist to help with thinking about what happens in next phases of migration work.

Scott - what about service providers who could implement Fedora? Similar to DSpace?

Erin - There are about 25 service providers for DSpace. It is expensive, though, and there are a limit to people who are able to code Fedora. Sharing information with service providers would be valuable to let them know where things are heading with Fedora.

Scott - also work with Lyrisis, consider possibilities, formal encouragement to develop service provider relationships.

David - good to connect with list of schools that need migrations and the service providers in some capacity. Samvera sites still running three all very similar to custom implementations. Migration will involve a lot of custom work. Question - to what degree can we build on and create tools that will help with custom shops. OCFL/6 will be path with custom stuff as it will be easier than getting to 4 or 5.

Erin - need to be deliberate about where we focus our efforts - some institutions will still need to manage their own migrations.

DAvid - front end - Islandora is uniform. custom shops - so many variations can't really make a tool.

Andrew - slides on migrating 3-4 and there are different categories they present in Fedora trainings ([see sample slides](#)): content migration, front-end migration, functionality migration (authorization controls, disseminators).

Este - we are planning to migration content and then look at user interface migration second. Like the categories for migration.

Scott - Tools to help with the migration in terms of how to plan for front end, such as how to map Solr from 3-4

David - initial migration, get content over - fairly low barrier, and not that bad in terms of moving to a new version of Fedora. Maybe create a Fedora migration guide?

Erin - great idea. Bridge to Hyku just put out their recommendations. Metadata clean up could wait?

Scott - take data model now and same bitstreams and put in Fedora 6. Maybe later tear things up and expose more linked data but not yet. This isn't that difficult.

David - don't have to convert all XML to RDF. It is not a requirement.

Migration tools

David - there are three that exist - [migration-utils](#), [FedoraMigrate](#), [Migrate_7x_claw](#) (based on [The Drupal Migrate framework](#)). Can we leverage what is there/update?

Erin - how heavy a lift to make migration-utils compatible for 5.0. A good starting point for the basic content migration. Would need some updating in terms of documentation.

Scott - documentation is a core are for our future development. Google doing 'summer of documentation' "[Google Season of Docs](#)" - offering stipends of up to \$6,000 for a 3 month documentation writer for technical documentation writers. Could be interesting for Fedora and Duraspace.

Erin - any feedback about the existing documentation about what should be addressed?

Scott - having documentation, and having it up to date. Having a good level of documentation is very helpful, it shows and makes it easy to develop and doing thing

Erin - helps to level the playing field.

Scott - writing good documentation is hard/a skill.

Erin - IMLS is focused on the US, documentation in English, should have documentation translated in other languages, and pursue funding.

David - need ongoing support for translation.

Este - could we start with the Fedora migration guide as a place to have multilingual support?

Tim - can we look at the survey to see what people need the most help with?

Erin - could we add content/metadata/ux migration into the survey?

Este - could we also mine Slack channels to see what kinds of questions come up?

David - other tools, fedora migrate gem - may not be as profitable to spend time on this one. More focused on migrating into specific Hydra/Sufia environment at the time.

Mike - specifically designed for folks using Sufia which is no longer a thing. Would take substantial work to have it function with Hyrax and that hits only a portion of the Samvera community, probably dead code at this point and not a good option for us to consider at this point.

David - migrate 7x in Islandora - Islandora working on this and it will do content migration.

Erin - are there things we can learn from this for what we should set up for Fedora migrate?

David - can configure in Drupal.

Andrew - what is Drupal doing that makes it so easy.?

David - Danny has done a YouTube video and it looks pretty straightforward. Would like to see what is working well and have it help in building out Fedora migration tool.

Este - concern about what happens to metadata in the Islandora migration.

Erin - would be good to talk to other types of staff like metadata folks.

Scott - we give people the option to look at their xml while building tools to get the data migrated.

David - migrationutils highest value to work on at this point.

Andrew - migrationutils reads Fedora 3 via API or what is on disk, and writes into Fedora via the API. Increasingly thinking of going from file system to file system. From tools he agrees, but there might be a higher level recommendation to write from disk to disk rather than API, which could be a new tool.

API documentation

David - summarizes API. Utility of API is already well known. You can migrate through the API but it is **slow**. REST, one object at a time. Have heard that the API ingest into 4 is so slow it is not useful for mass migration. Could take 6 months to move data. NLM projected at least six months.

Scott - stumbling block. Interest in performance and scalability. Finding a bottleneck with the triplestore in Fedora 3. Not a Fedora problem necessarily.

Andrew - use half of the migration tool - reads from Fedora 3 and writes to Fedora 6 in OCFL.

David - the stuff is on disk. Suggestion that Fedora 6 could write to comply with objects on disk rather than other way around where objects conform to what is acceptable to the application. Could still use API, but could just write to disk in OCFL compliant format and that would be faster.

Erin - are we going to recommend people wait until Fedora 6?

David - depends on the amount of content.

Andrew - 6 is resolving issues in Fedora 4 and 5. Fedora 3 has flat hierarchy. Fedora 4 and 5 requires to change urls.. in Fedora 6 can keep urls and keep flat hierarchy, and have no modeshape.

Scott - would almost be two different migration utilities from 3 - 4/5 and 3-6.

Erin - grant focus is Fedora 3. Fedora 6 is well underway. Don't know the timeline yet. List of recommendations to publish with grant project - who to talk to about at what point server conversations with IT about last date of moving. Two rare instances where institutions who need to migrate or they have to turn it off.

Scott - this is a coming issue. Java 11 is coming at his institution, and Fedora 3 won't work on Java 11.

Mike - have had some conversations. No one says will stop supporting Java 8. Threat of something tanking and having a problem is getting greater.

Scott - no bug fix for end of life for Java 8. They use free SDK. Part of reason to migrate to 11 is to use the open JDK. Other issue with Fedora 3 - a lot of libraries and jars in it that are old and have bugs and security flaws that aren't being updated, and will get pushback from security admins.

Tim - in charge of everything from wall plate out - they aren't worried from pressure from IT at this point.

Erin - would like to send out an email to the whole advisory board about decommissioning Fedora for security/support perspectives, and who is applying the pressure.

David - this is a major problem. It could be a problem for us.

Erin - means the time pressure for 6 is even greater. This is interesting...

Tim - sees this in the espoused mission of DPN - to make sure that content wasn't lost. We are building a case that the lost may not be the preservation strategy, but people aren't resourced to get their stuff out of an aging infrastructure. Not just the technology uplift from 3-6, but also shop it around as a crisis or threat to loss of content. The more we focus on that, the more evidence that it is a cross institution national issue, not just a technology issue.

Erin - this seems like a compelling argument to IMLS.

Scott - another key for Fedora 6 is the OCFL, which speaks to preservation.

Este - we will go to 6, but not 4 or 5, but we need to have it go well.

Erin - we need to support the first couple migrations to 6, with site visits to make sure it is good.

Scott - Bitcurator consortium - has a deal if you join for 6 months, you get three days of support. They walk you through a process - targeted at a very concrete level for institutions not well resources or technically in depth. Can put in contact with some folks from Bitcurator. Uses video calls, not on site.

Erin - a program institutions could apply for - very interesting idea beyond picking a few institutions to start with.

OCFL

David - helps with speed issues with migration and batch ingest.

Scott - solving not just a migration problem as Tim said, but a preservation problem.

Erin - expanding on the risks to the content in terms of the deprecated applications.

Survey

Tim: additional question: what most do you need help with? (UX, content migration, metadata, etc.)

Erin: what skills are in short supply at your institution?

Este: what are you most concerned about, worried about? (time, number of objects, metadata)

Suggestion: what are barriers? check all that apply

David: keep survey short, 5 - 10 minutes; but an additional question about what you need help with is good

Este: open ended? Erin: take time and thought to respond to

David: compromise: options + Other free-text field

Erin: add the question below barriers question

Erin: in terms of help: how do we classify kinds of help?

David: let respondents determine the answers

Este: rephrase: "what would be helpful?"

David, et al: UX help, tooling, documentation, consultation, training, service provider contracting, metadata

Scott: helpful to get responses, even if we can't help with everything

Survey Timing and distribution

Erin: release in a week, keep open 2/3 weeks; send two reminders; first solicit personal contacts, then to open lists

direct contact distribution list: TODO for group: please brainstorm individuals who can fill out survey (5 - 10)

add to Excel spreadsheet

Este, David: emphasize that anyone, any role, can answer, don't need specialized knowledge, not a team effort

Erin: will send out list of institutions for initial grant

Erin: timing will work well with when intern come on board

Erin: do we need a webinar (as promised in the grant proposal) before the survey, to orient people to the survey, communicate grant progress?

Este: yes, more informational, no need for heavy prep

Andrew: and record it

Erin: send it out next week, open for three weeks, have informational call with second or third reminder

Mike: who is audience? Class of respondents who have migrated away, who may have resources, but choose not to migrate?

Erin/David: hopefully, both

Erin: hopefully, 80 -120 respondents

Scott: seems high

Este: stress that we want more than one person at a given institution

Survey preamble

Should take 10 minutes

Andrew: and specify number of questions

Erin: also stress that no prior knowledge or expertise required

Erin: preservation focus in preamble?

Scott: yes: blah blah blah (will be wordsmithed)

Este: upgrading vs. migrating?

Erin, David: barriers to migrating (not upgrading)

(some wordsmithing)

Scott: privacy/anonymization boilerplate? what will we do with the data?

Erin: we solicit institutional affiliation, but not names 6th

Scott: add some boilerplate about how data will be collected, used, and reported

Tim: get people to click: three or four bullet points at top?

Erin: subject line to induce people to open the email?

Scott: include preservation?

Erin/Este: too broad...

Este/Andrew: short and punchy

Erin: Barriers to Fedora migration: quick survey

(update preamble to same subject/title)

Survey distribution

Erin: send out by Wednesday?

Advisory group: add names to list by Wednesday (Apr. 17th); close survey on May 8th

Scott: markers for phone call, first reminder, second reminder?

Erin: first reminder: April 24th, David Wilcox will mention at Samvera presentation

Phone call: April 25th or 26th

Last reminder: May 6th

Erin: survey preamble as template for email

Planning Next Steps

- Need to follow up with community stakeholders and advisory board members with the survey results.
 - Do the advisory board consultations need to be individual or can they be done as a group?
 - We can schedule an advisory board call a week after the survey closes to discuss the results
 - Monday, May 13 at 11am ET
 - Most of the results will be multiple choice/checkboxes so it shouldn't take long to aggregate and report
 - This is another form of primary research
- Should we collect names and email addresses as an optional field?
 - Could be useful for secondary consultation - e.g. invite people to a Zoom call to ask follow up questions
 - Add an optional field to the end of the form and update the preamble with regard to privacy
 - Draft questions and follow up with select respondents on a Zoom call
 - Make sure form logic works so that all participants are invited to provide contact details
- If we finalize our work by July 1 we can just submit, otherwise we'll need to transfer to Lyrasis which can take some effort

Final Report

- Where to publish survey results
 - Link excel spreadsheet to the wiki/website
 - Scott will check to see if we can deposit to UW IR
 - We should mint a public PID (ARK, Handle, DOI, etc.)
- What should be included in the final report?
 - Summary of this morning's discussion topics
 - Clear indication of what is needed for a follow-on grant (docs, tooling, etc.)
 - Recommendations, prioritized tasks to present to Fedora governance leading to concrete actions
 - Not a SWOT analysis, but a statement of the risks involved in not migrating away from Fedora 3.x
 - Clarity about what DuraSpace/Lyrasis should do vs. what other stakeholders should do
 - This will help with communication with other groups/service providers
 - Prioritization based on need and impact
 - Description of where this fits in with other DuraSpace/Lyrasis priorities (e.g. OCFL)
 - Consider SOAR (Strengths, Opportunities, Aspirations, Risks) analysis
 - Confirmation that people in the community aren't alone, DuraSpace/Lyrasis is on top of things
 - Beyond just migrating from 3.x to 6.x, thinking about future versions of Fedora
- Should we check in like this every few years?
 - What problems are you facing, what do you need help with
 - This helps us get beyond our current governance structure, which heavily favours institutions that are able to pay membership fees
- The results of this effort could be useful to other communities as well (e.g. DSpace)
 - The process could be useful to share as well
- Dissemination
 - Presentation accepted at OR
 - The group will put a proposal into DLF
 - Islandoracon, Samvera Connect, CNI

- Organize a webinar in the Fall
 - Publish a paper in an open access journal (e.g. Code4Lib)
 - Include in workshops and camps
 - Couple messaging with release of Fedora 6
- Next Steps in the grant process
 - This is a planning grant, implementation grants typically follow
 - IMLS will put out a call for proposals in September
 - Support for a second phase would probably fall into 2020
 - What could a follow-on grant help fund in this timeline?
 - Hard to know before we see the results
 - Identify priorities in the report without assuming any grant funding, but indicate where funding would help accelerate these efforts
 - We want to be working on migration tooling as soon as possible
 - The grant could help fund migration guide, site visits, training
 - Probably target \$150K
 - Revisit this once we have the results of the survey on the May 13 meeting

Actions

- [David Wilcox](#) add "What would help" question to the survey
- [David Wilcox](#) Add option name and email address fields to the end of the survey
- [David Wilcox](#) Send out survey to mailing lists on April 17
- All: Add 5 people to the list of contacts for personal survey invitations
- All: Draft proposals for DLF, Islandora Con, Samvera Connect, CNI