

2012-03-05 DfR Technical Meeting

Regular Attendees

- Andrew
- Bill
- Brad ★
- Chris
- Dan
- Jonathan

General

- Call In To: [Free Conference Call HD - DuraCloud Line](#)
Please note that we will be using the Free Conference Call HD line for this call. Information about calling into this line is available from the link above.
- ★ - Indicates who will be taking minutes

Discussion Topics

[Planning Board](#)

[Task Board](#)

[DfR Planning and Estimating - 0.1](#)

Topic	Discussion Leader
Updates to Support OCS - Add on Tasks for Iteration 1	Chris
Lessons Learned from Islandora Demo	Dan
Planning and Task Boards - Review and Building the Backlog	Dan
Synchronization (Scope DFR-23)	Bill
Project and Partners	Jonathan

Actions from last meeting

[Last meeting](#)

Last Weeks Actions:

Action Item	Assignee	Status
Rephrase DfR 1 to better fit future iterations	Dan	
Rephrase DfR 74 wrt Fedora	Dan and Chris	
Better conceptualize accounts for DfR	Team	Postponed
Remote vs Web UI for Synch Tooling	Bill	
Better describe feasible scope for this year in automated metatdata collection	Team	
Tie some of the DuraCloud jira items to DfR stories	Bill + Andrew	✓
Set up implementation methodology docs	Chris	✓

[Related Information](#)

[DfR Software Next Steps - Jan 2012](#)

[2012-01-03 - Architecture Meeting, Temple University](#)

Status

- Jonathan
- Andrew
- Bill
 - In progress: [DFR-74](#).

- First step, creating wrapper [DuraBoss application](#) is complete
 - Next step, create service handlers
- Brad
- Chris
 - Finished tasks for [DFR-11](#) and got a working [end-to-end OCS ingest & delete test](#) going between DuraCloud & Fedora.
 - While doing DFR-11 tasks, identified a couple stories (using "E" type datastreams and persisting MD5 checksum in Fedora) that should be written for an upcoming iteration. Both of these require outstanding Fedora issues to be addressed.
 - Started [DfR Development Resources](#) with various info/links important to DfR developers, including coding guidelines which should be reviewed by others dev-types (try out the intelliJ settings attached to that page).
- Dan
- Jonathan

Minutes

- Chris on OCS - Two issues discovered::
 - Getting round trip to Fedora working was goal - and is completed.
 - Issue: Can't turn off checksum checking in Fedora.
 - Issue: Authenticating to retrieve content - need to add support to Fedora; implementation exists but needs to be progressed.
 - Jira issues TBD
- Note that it is important to set our scope correctly to deliver something useful.
- Islandora / Smithsonian Demo impressions:
 - Dan: Smart interface that avoids generic hand coding of items; not yet answered is to get over the tipping point of what researchers will use.
 - Jonathan: App needs to be very user friendly to succeed.
 - Dan: Note that you can also see the same data in a traditional website out of the same tool. Can customize the views to provide branding / institutional identification, etc.
 - Bill: Quite flexible in setting up the data structure. Not sure what it would look like as a defined set of concepts that the researcher understands. Expect that any given domain will have it's own set of concepts that should be emphasised. Where does it fit in the flow of the file upload interface?
 - Chris: How would we integrate into DFR at this point? Lots of value if you incent the researcher to enter data into a webform ala Islandora demo.
 - Brad: Can we keep track of items the research has seen or hasn't? If the life cycle (in somewhat sequential order) is:
 - Identify item to be uploaded
 - Upload the item
 - Specify rules for auto tagging
 - Review the tagging and/or enter item specific details
 - View the system default view of the data (set)
 - Define and use better views of the data
 - Transfer the data to someone else
 - Dan: We need to get the data, and then make sure it is well characterized and want as automated a fashion as possible. What is the forms are prepopulated, for example.
 - Dan / Jonathan - it will be easy to see the uploaded items in Islandora
 - Jonathan / Brad : So do we build those contact points with the user as part of Islandora or our version of Islandora? Maybe, but we want to find the right implementation.
 - Dan: Mark is already starting to think about moving components down the stack. We are in the middle space, working with both the GUI and the low level. Chris has been building the OCS; it is a generic processing pipeline; even if Islandora started the creation of a Fedora Object, you could still use the OCS there. So not an enhanced version of Islandora, but working with that reusable pieces of OCS.
 - Brad: Yes, both Jonathan's idea of reusing Islandora's interface for control of DFR actions and Dan's OCS pipeline for underlying tech makes sense.
 - Jonathan: Does that mean we need Drupal skills on the UI team?
 - Dan: Also need data extraction experts (Harvard, file characterization and classifying things?). Tech from search engines might apply (geolocation, other)
 - Dan/Jonathan - ideally we want "free" resources, but might need to make a few strategic time purchases to get integrations.
 - Bill: See how far we get with the FITS tooling from harvard (JOVE + DROID, etc) before bringing someone in.
 - Dan: Other tech to bring in would be geolocation; extract. Islandora has integrated a search engine; can we add SOLR (plus the existing triple store); full search index plus geolocation tagging.
 - Brad/Dan: So the goal is to get a happy path that pulls the geoloc data and puts in in a form that existing Islandora tech can visualize it.
 - Bill: Should we look and see what visualizations are already available and use this to guide what we try to extract?
 - Bill: Next steps with smithsonian?
- On Sync and DFR-23:
 - Dan: What UI to use?
 - Bill: Next step is to distinguish what sync needs DFR has that aren't the same as DuraCloud
 - Bill: Would like to go for a locally deployed webapp - similar user experience. Local means a webapp on their local machine.
 - Dan: Can we scope how much setup and additional data we will limit ourselves to right now? Are we going to do directories, how do we specify? (we are doing now).
 - Chris: What beyond current sync abilities do we want?
 - Multiple users? Options -> separate spaces, separate folders, How does project in smithsonian map to the directories in DuraCloud?
 - Dan: This issue to be revisited later. Please keep sharing thoughts on the design.
- Jonathan on UI Resources:
 - Charles might or might not be available; will know later this month.
 - Still looking for resources at MIT.
 - Dan: Jonathan, should you talk to the FITS folks? (Andrew has had conversations in the past)

Actions

- Regular status checks of Brad/Jonathan/Thorny/Mark to keep up to date.
- Dan takes point on talking with Paul Pound to keep in syncon implementations.

