

# 2019-06-19 Specification meeting

## Date

13 May 2019

<https://duraspace.zoom.us/j/593874187>

Find your local number: <https://zoom.us/j/X8NrH>

## Attendees

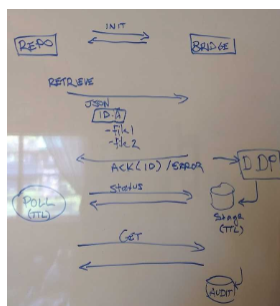
- Michael Ritter
- Bill Branan
- tamsin johnson

## Regrets

- Sibyl Schaefer

## Goals

## Diagram



## Discussion items

Time	Item	Who	Notes
	Current Work	Tom	<ul style="list-style-type: none"><li>• Created 3 pages for Hyrax specs/workflows<ul style="list-style-type: none"><li>◦ <a href="#">OTM Hyrax Preservation Policy Specification</a></li><li>◦ <a href="#">OTM Hyrax Preservation Workflow Specification</a></li><li>◦ <a href="#">OTM Repository Gateway Specification</a></li></ul></li><li>• Gateway is the main interaction with the Bridge rather than the repository. From Bridge pov it doesn't worry about the repository.</li></ul>

	OTM Bridge Spec	<p>Review of Bridge API spec: <a href="#">OTM Bridge API Specification</a></p> <ul style="list-style-type: none"> <li>Assumptions <ul style="list-style-type: none"> <li>Bridge would have a user registered by the DDP in order for the gateway to communicate, can discuss authentication later on</li> <li>file ids are unique per account – assuming that the gateway/repository has some way of maintaining its own uniqueness. file ids can collide between accounts.</li> <li>some type of authentication on endpoints - can consider basic auth to start, potentially using tokens instead <ul style="list-style-type: none"> <li>some endpoints might not need auth depending on where they are (e.g. getting files from the gateway)</li> </ul> </li> </ul> </li> <li>Register <ul style="list-style-type: none"> <li>what type of authentication to use - should we do something akin to aws with signed tokens?</li> <li>basic auth to start?</li> <li>otm-url to gateway or repository? <ul style="list-style-type: none"> <li>gateway</li> </ul> </li> </ul> </li> <li>Version <ul style="list-style-type: none"> <li>specification API or application API? <ul style="list-style-type: none"> <li>s...pecification?</li> </ul> </li> </ul> </li> <li>Deposit <ul style="list-style-type: none"> <li>is the deposit-id necessary? <ul style="list-style-type: none"> <li>no, the bridge could return it in a response. moves responsibility of uniqueness to the bridge + simplifies request body</li> </ul> </li> <li>have been avoiding the question of what constitutes a file, but must ask: what is a file? <ul style="list-style-type: none"> <li>a bytestream</li> </ul> </li> <li>are files being grouped in any way? <ul style="list-style-type: none"> <li>grouping implies we need to store extra information associated with the files and be able to track that</li> <li>current spec is simple - just push file identifiers</li> <li>generally have been trying to avoid describing any grouping abstraction (collection, work, bag, etc), might be best to continue to avoid it</li> <li>possibly add idea of groupings later</li> </ul> </li> <li>checksum-type is defined at the top level - what about per file? <ul style="list-style-type: none"> <li>will expect that all deposits use the same checksum, but can see scenarios where the other might exist</li> <li>I don't remember where we landed, probably keep this for now and add on later</li> </ul> </li> <li>add endpoint for supported checkpoints? <ul style="list-style-type: none"> <li>maybe extend the version api with more information</li> </ul> </li> </ul> </li> <li>List Deposits <ul style="list-style-type: none"> <li>on lists deposits in progress, intended to provoke thoughts</li> <li>deposits are seen as an ephemeral resource, not something we will be keeping forever</li> <li>failed deposits would persist until aborted or restarted</li> <li>this is something we should continue to dig into and think about</li> </ul> </li> <li>Deposit Status</li> <li>Abort Deposit <ul style="list-style-type: none"> <li>only valid in certain states</li> <li>need to discuss guarantees from the bridge about what actions it will take depending on where a deposit is</li> </ul> </li> <li>Restart Deposit <ul style="list-style-type: none"> <li>similar to abort, just needs to be livened up a bit</li> </ul> </li> <li>Delete Content <ul style="list-style-type: none"> <li>are we doing a delete or a purge? <ul style="list-style-type: none"> <li>what is the difference?</li> <li>we can update the endpoint to delete for now, reserve purge for a future version when people have strong opinions about how they want to remove content</li> </ul> </li> <li>what are the error responses/conditions? <ul style="list-style-type: none"> <li>checksums not matching, files not existing</li> <li>how much work do we expect the bridge to do on operations <ul style="list-style-type: none"> <li>transferring 10 1tb files could mean certain operations are expensive</li> <li>even procuring storage</li> <li>needs more discussion, maybe insight from user stories</li> </ul> </li> <li>when a single file fails, does that fail the entire deletion? <ul style="list-style-type: none"> <li>I don't remember</li> </ul> </li> </ul> </li> <li>what is the url referring to? <ul style="list-style-type: none"> <li>do we need the url? no</li> <li>do we need the url in the deposit? no - just need to ensure that the gateway has a well defined way of hosting by file identifiers</li> </ul> </li> </ul> </li> <li>Delete Status <ul style="list-style-type: none"> <li>needs discussion on what the response body is</li> <li>give per file information?</li> </ul> </li> <li>Get Audit History <ul style="list-style-type: none"> <li>file-id not guaranteed to be unique across accounts, maybe we shouldn't use it <ul style="list-style-type: none"> <li>could update audit history endpoint to be POST with a request body indicating what files, events you want to receive</li> <li>allows for good extensibility later on</li> <li>aligns well with to other endpoints</li> </ul> </li> <li>should this return a response body or an audit-id which the audit events can be received from? <ul style="list-style-type: none"> <li>depending on how much is being asked the operation could take some time. could easily trigger timeouts, etc.</li> <li>if we push to an audit-id, we need to have an idea of how long that resource would last <ul style="list-style-type: none"> <li>other events are ephemeral, is this too?</li> <li>need additional endpoint if we do this</li> </ul> </li> </ul> </li> </ul> </li> <li>Restore Content <ul style="list-style-type: none"> <li>do we need the checksum? <ul style="list-style-type: none"> <li>maybe optional</li> <li>gateway might not have any idea of what the checksum is</li> </ul> </li> <li>need additional endpoint to say where the file is located <ul style="list-style-type: none"> <li>restore/{restore-id}/{file-id} <ul style="list-style-type: none"> <li>if restore-id is unique, should not need to worry about collisions</li> </ul> </li> </ul> </li> </ul> </li> <li>Restore Status <ul style="list-style-type: none"> <li>what even is status</li> </ul> </li> </ul>
--	-----------------	--

	Next steps	<p>Tom - work on Repository Gateway Sepc</p> <p>Bill - updates to <a href="#">OTM Bridge API Specification</a> from discussion</p> <p>Mike - work on DDP/Bridge interaction + put on wiki</p> <p>All - deep thinking about life, liberty, and the pursuit of storing data forever</p>
--	------------	---

## Action items

-