

DevMtg 2019-03-06

Developers Meeting on Weds, March 03, 2019

Today's Meeting Times



- DSpace Developers Meeting / Backlog Hour: 15:00 UTC in [#duraspace IRC](#) or [#dev-mtg Slack channel](#) (these two channels sync all conversations)

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- Below are the last published dates. This editor is unsure of the next ones. Some recent meetings have been cancelled due to various conflicts.
- [DSpace 7 Working Group \(2016-2023\)](#): Next meeting is Thurs, Feb 14 at 15:00 UTC. Agenda: [2019-02-14 DSpace 7 Working Group Meeting](#)
- [DSpace 7 Entities Working Group \(2018-19\)](#): Next meeting on Tues, Feb 19 at 16:00 UTC
 - Last meeting notes at [2019-02-05 DSpace 7 Entities WG Meeting](#)
- [DSpace Developer Show and Tell Meetings](#): On hold until interesting topics arise.

Discussion Topics

If you have a topic you'd like to have added to the agenda, please just add it.

1. (Ongoing Topic) [DSpace 7 Status Updates](#) for this week (from [DSpace 7 Working Group \(2016-2023\)](#))
2. (Ongoing Topic) [DSpace 6.x Status Updates](#) for this week
 - a. 6.4 will surely happen at some point, but no definitive plan or schedule at this time. Please continue to help move forward / merge PRs into the dspace-6.x branch, and we can continue to monitor when a 6.4 release makes sense.
3. [Upgrading Solr Server for DSpace \(Mark H. Wood\)](#)
 - a. PR <https://github.com/DSpace/DSpace/pull/2058>
 - b. Docker configuration for external Solr
 - i. <https://github.com/Georgetown-University-Libraries/DSpace/commit/7115173d61776dd2455690518f5c9809cd0f28d4>
 1. The Dockerfile creates a new solr instance with 4 cores. It then overlays the schema and config changes in PR 2058.
 2. I attempted to create my branch so that I could create a PR back to Mark's branch, but some other changes from master seem to be showing up if I create a PR.
 - ii. This will need a small change to our docker compose files to invoke the external solr service. <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/79>
4. [DSpace Backend as One Webapp \(Tim Donohue\)](#)
 - a. PR: <https://github.com/DSpace/DSpace/pull/2265> (PR is in a reviewable state. SWORDv1 and SWORDv2 are merged into "Spring REST" webapp, with basic Integration Tests to prove both work)
5. [DSpace Docker and Cloud Deployment Goals \(old\) \(Terrence W Brady\)](#)
 - a. Add Docker build/push to Travis
 - i. This make sense to consider after 2307 is merged
 - ii. <https://github.com/DSpace/DSpace/pull/2308>
6. Brainstorms / ideas (Any quick updates to report?)
 - a. (On Hold, pending Steering/Leadership approval) Follow-up on "DSpace Top GitHub Contributors" site (Tim Donohue): <https://tdonohue.github.io/top-contributors/>
 - b. [Bulk Operations Support Enhancements](#) (from Mark H. Wood)
 - c. [Curation System Needs](#) (from Terrence W Brady)
7. Tickets, Pull Requests or Email threads/discussions requiring more attention? (Please feel free to add any you wish to discuss under this topic)
 - a. Quick Win PRs: <https://github.com/DSpace/DSpace/pulls?q=is%3Aopen+review%3Aapproved+label%3A%22quick+win%22>

Tabled Topics

These topics are ones we've touched on in the past and likely need to revisit (with other interested parties). If a topic below is of interest to you, say something and we'll promote it to an agenda topic!

1. Management of database connections for DSpace going forward (7.0 and beyond). What behavior is ideal? Also see notes at [DSpace Database Access](#)
 - a. In DSpace 5, each "Context" established a new DB connection. Context then committed or aborted the connection after it was done (based on results of that request). Context could also be shared between methods if a single transaction needed to perform actions across multiple methods.
 - b. In DSpace 6, Hibernate manages the DB connection pool. Each **thread** grabs a Connection from the pool. This means two Context objects could use the same Connection (if they are in the same thread). In other words, code can no longer assume each `new Context()` is treated as a new database transaction.

- i. Should we be making use of `SessionFactory.openSession()` for READ-ONLY Contexts (or any change of Context state) to ensure we are creating a new `Connection` (and not simply modifying the state of an existing one)? Currently we always use `SessionFactory.getCurrentSession()` in `HibernateDBConnection`, which doesn't guarantee a new connection: https://github.com/DSpace/DSpace/blob/dspace-6_x/dspace-api/src/main/java/org/dspace/core/HibernateDBConnection.java
- c. Bulk operations, such as loading batches of items or doing mass updates, have another issue: transaction size and lifetime. Operating on 1 000 000 items in a single transaction can cause enormous cache bloat, or even exhaust the heap.
 - i. Bulk loading should be broken down by committing a modestly-sized batch and opening a new transaction at frequent intervals. (A consequence of this design is that the operation must leave enough information to restart it without re-adding work already committed, should the operation fail or be prematurely terminated by the user. The SAF importer is a good example.)
 - ii. Mass updates need two different transaction lifetimes: a query which generates the list of objects on which to operate, which lasts throughout the update; and the update queries, which should be committed frequently as above. This requires *two* transactions, so that the updates can be committed without ending the long-running query that tells us what to update.

Ticket Summaries

1. Help us test / code review! These are tickets needing code review/testing and flagged for a future release (ordered by release & priority)

key	summary	type	created	updated	assignee	reporter	priority	status	fixversions
Unable to locate Jira server for this macro. It may be due to Application Link configuration.									

2. Newly created tickets this week:

key	summary	type	created	assignee	reporter	priority	status
Unable to locate Jira server for this macro. It may be due to Application Link configuration.							

3. Old, unresolved tickets with activity this week:

key	summary	type	created	updated	assignee	reporter	priority	status
Unable to locate Jira server for this macro. It may be due to Application Link configuration.								

4. Tickets resolved this week:

key	summary	type	created	assignee	reporter	priority	status	resolution
Unable to locate Jira server for this macro. It may be due to Application Link configuration.								

5. Tickets requiring review. This is the JIRA Backlog of "Received" tickets:

key	summary	type	created	updated	assignee	reporter	priority
Unable to locate Jira server for this macro. It may be due to Application Link configuration.							

Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Meeting Notes

Meeting Transcript

Log from #dev-mtg Slack (All times are EST)

Mark Wood [10:00]

Welcome to the weekly developer meeting, all!
There's a link to the agenda above.

Terry Brady [10:01]

Hello

Mark Wood [10:01]

Shall we see who else is here?
It seems that we are three today.

Alexander Sulfrian [10:02]

Hi

Terry Brady [10:03]

The Docker for Repository Managers webinar went very well yesterday. <https://duraspace.org/webinar/> (edited)

Mark Wood [10:03]

Quick reminders: the DSpace 7 and Entities working groups are still working, and do continue with scheduled meetings, although some recent meetings have been cancelled due to conflicts.
Hi @sulfrian, we are four now.
Anything else to say about the webinar, @terrywbrady?

Terry Brady [10:05]

We had 55 people and a dozen great questions at the end.
One attendee mentioned that they were able to work through the sample exercises while we were on the call. I hope to hear more from other attendees over the next week.

Mark Wood [10:07]

That does sound as though it is working well.
Since we're talking Docker anyway, shall we skip ahead and see if there is anything else to discuss on that topic today?

Terry Brady [10:09]

Sounds good. I merged a PR yesterday on master that allows update sequences to run via the database command. I plan to port that to 5x and 6x. @tdonohue cautioned that it might not port as easily to 4x.

Mark Wood [10:10]

I see that several of the PR links I copied are to merged PRs. I'll update the lists after the meeting.

Terry Brady [10:11]

The Docker image we have for Oracle is handy. I have reached out to Atmire to see if they would want to chat about support for the dspace-oracle image. It would be great to make that a pluggable option into the existing docker compose files.
@tdonohue shared a link to the entities wg dataset. I plan to explore how that could be used as a sample AIP loaded dataset for Docker images.
That is it for me on Docker.

Mark Wood [10:12]

Comments from others?
OK, we'll move on. Back to the top of the list: I have no status updates for 6_x or 7_x.

Tim Donohue [10:14]

Sidenote for @terrywbrady (sorry, in another meeting, so just lurking): We may not be able to create AIPs of

the Entities dataset -- as all aspects of Entities (especially relationships between them) are not yet represented in AIPs. It's worth a try, but not sure if it'll work

Mark Wood [10:15]

Comments from others on those?

None, it seems. On the Solr upgrade: <https://github.com/DSpace/DSpace/pull/2058> has had another review, and has an open change request. What does it need to make it approvable?

I will spend some time today on seeing what is needed for upgrading an older DSpace instance and moving its cores to a separate Solr install.

Terry Brady [10:19]

Sorry that I did not update my review. I think I was waiting to see if you needed another test from me.

Mark Wood [10:20]

I mainly need to have addressed your concerns so that the request can be closed.

My thanks to everyone who has reviewed it.

Terry Brady [10:21]

You had addressed all but the statistics issue before my prior review. I have not yet re-validated your fix to the statistics object.

I'll do another quick look at the code.

Mark Wood [10:22]

OK, thank you!

Other comments on the Solr upgrade?

OK, moving on: DSpace Backend as one webapp. No updates that I know of, and Tim is unavailable at the moment. Reviews would be helpful, I am sure. (I should review it.)

Tim Donohue [10:25]

A quick update, the PR is fully ready for review: <https://github.com/DSpace/DSpace/pull/2265>

Mark Wood [10:25]

Ah, thank you!

Tim Donohue [10:25]

see latest comment there for more details

James Creel [10:26]

I've got to run to another meeting in 5 minutes - I might jump in and give an update on support for deleted objects in OAI.

Mark Wood [10:26]

OK, please do.

James Creel [10:27]

So the librarian stakeholders here are keen to get "persistent" support for deletions in OAI. Currently DSpace is supposed to do this for withdrawn items but not for deleted ones which have not tombstone.

Now, I tested the withdrawal use case on a DSpace 6.3 build and it wasn't working for me, so there may be a bug. I'll investigate further to confirm.

Supporting it for actually deleted items would involve storing some persistent tombstone. Would others find this objectionable in principle?

Probably would involve change to db schema

Mark Wood [10:29]

How does one actually remove an item so that it is actually gone, no longer in evidence? Someone is going to want that.

James Creel [10:29]

Right - I agree. But it fundamentally conflicts with achieving "persistent" support for deletions in OAI.

Mark Wood [10:29]

A "tombstone record" could be an Item with no content streams.

James Creel [10:30]

This drives people crazy with Fedora, how you have to delete the tombstones

Perhaps it could be configurable

Anyway, I have got to run. I'll jump back on this channel or dev channel in just a few minutes. Want to mention that we will be doing a DSpace sprint here at TAMU starting March 25.

Mark Wood [10:31]

We probably need some new terms. To me, something deleted is no longer in existence. We need a term for "no

longer offered but you can still know that we had it."

Alexander Sulfrian [10:32]

This does work for withdrawn items.

Mark Wood [10:32]

I would have said that "withdrawn" denotes that state.

I should look and see whether a withdrawn Item can be recovered, i.e. un-withdrawn. That would be yet another state.

Alexander Sulfrian [10:34]

This is a withdrawn item in OAI: https://refubium.fu-berlin.de/oai/request?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:refubium.fu-berlin.de:fub188/22673

Terry Brady [10:34]

Pasted image at 2019-03-06, 7:34 AM

Mark Wood [10:36]

So possibly desired states of an item: submitted, not yet available; in-archive; withdrawn (we have it but you can't get it); deleted (we had it but it is gone); purged (no longer detectable).

Not to mention embargoed, private....

Terry Brady [10:37]

Do we consider the message that displays for a withdrawn item to be a tombstone?

Pasted image at 2019-03-06, 7:37 AM

Mark Wood [10:37]

For completeness: submission in development.

Perhaps. I'm still unclear on whether a tombstone belongs to something withheld from view, something actually no longer in the archive, or both.

The use of the word makes me think it means dead-and-gone, no longer in the archive.

Terry Brady [10:40]

These states would be nice to document clearly for end-users. If I remember correctly, there are 3 booleans in the item table and a couple of them seemed redundant to me until I had used DSpace for several years.

Mark Wood [10:41]

That probably means that we should represent them as states, as in positions in a state diagram, rather than separate booleans.

But we also need a clear understanding of what the community wants, so that we are implementing *useful* states.

Terry Brady [10:42]

The most recent discovery to me was the representation of item templates in the item table.

Mark Wood [10:42]

That probably *is* a case for a boolean.

Hm, what we have is in_archive, withdrawn, discoverable.

Alexander Sulfrian [10:45]

First step could be to create a state diagram of the current possibilities and then discuss changes afterwards.

Mark Wood [10:45]

That seems good.

We then need more information from librarians on what they want to do.

Terry Brady [10:47]

I imagine that any changes to this behavior should be introduced in a major release rather than in a point release... unless there is no clear sense of expected behavior.

Mark Wood [10:47]

Yes.

But, good documentation of what DSpace does now could be added anytime.

Terry Brady [10:49]

Definitely. Where (within the wiki) is the best location for this documentation?

Mark Wood [10:49]

Good question....

Terry Brady [10:49]

It would be nice to document the anticipate behavior of all 8 (2^3) states.

Mark Wood [10:49]

Using DSpace | Items and Metadata?

Terry Brady [10:50]

That sounds like a smart place.

I need to jump away for 5 min. I'll loop back if you all have any questions for m.e

Mark Wood [10:50]

OK.

I don't want to interrupt discussion, but we have about 9 minutes left, so if there are other topics we should bring up, this is about the time to request the floor.

Otherwise we can continue on tombstones, or wrap up early.

I will mention that the Entities work brings new possibilities, because Item will now have a "type". So, for example, "template Item" could be a type.

Have we any further discussion today?

Terry Brady [10:55]

I'm back.

Mark Wood [10:55]

Does anyone want to take on drawing the Item state diagram?

Alexander Sulfrin [10:55]

I can try to draft something.

Mark Wood [10:56]

Thank you.

Should we close the meeting, or are there any final topics?

Tim Donohue [10:57]

Quick sidenote. If you haven't seen it recently, there's a list of DSpace item states here: <https://wiki.duraspace.org/display/DSDOC6x/DSpace+Item+State+Definitions>

Mark Wood [10:57]

Aha!

Tim Donohue [10:57]

So, any diagram should go there. Updates welcome too

Terry Brady [10:58]

A table showing the combinations of the 3 flags would be good as well.

Mark Wood [10:58]

We should make certain that there is a link to that from somewhere in Using DSpace.

Our time is ended. Unless there are objections, I'll close the meeting now.

Terry Brady [11:00]

Have a good week!

Mark Wood [11:00]

Thank you all. Meeting adjourned.