

DevMtg 2019-01-23

Developers Meeting on Weds, January 23, 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

- [DSpace 7 Working Group \(2016-2023\)](#): Next meeting is Thurs, Jan 24 at 15:00 UTC
- [DSpace 7 Entities Working Group \(2018-19\)](#): Next meeting on Tues, Feb 5 at 16:00 UTC
 - Last meeting notes at [2019-01-22 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](#) (Any status updates?)
 - a. PR <https://github.com/DSpace/DSpace/pull/2058>
4. [DSpace Docker and Cloud Deployment Goals \(old\)](#) (Terrence W Brady)
 - a. Simplify invocation by using multiple fragments, auto load content on startup
 - i. <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/68>
 - ii. Summary page: https://github.com/DSpace-Labs/DSpace-Docker-Images/blob/master/helper_cmds/docker-compose-files/dspace-compose/ComposeFiles.md
 - b. Speed up Docker builds
 - i. <https://github.com/DSpace/DSpace/pull/2307>
 - c. Add Docker build/push to Travis
 - i. This make sense to consider after 2307 is merged
 - ii. <https://github.com/DSpace/DSpace/pull/2308>
5. 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 Mark H. Wood)
 - i. PR 2180 improves reporting. Ready for review.
6. 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/master/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.							

Meeting Notes

Meeting Transcript

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

Tim Donohue [9:01 AM]

@here: It's DSpace DevMtg time. Agenda is at: <https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-01-23>

Terry Brady [9:01 AM]

hello

Mark Wood [9:01 AM]

Howdy

Tim Donohue [9:01 AM]

As mentioned in #dev, I haven't received any new topics to add to the agenda, so it's just an updated agenda of last week's topics. If we run out of things to talk about though, we can always end early :wink:
And thanks for already starting the roll call!

James Creel [9:02 AM]

Hi everybody

Tim Donohue [9:02 AM]

So, let's go ahead and get started here. Welcome all

James Creel [9:02 AM]

I have to jet at 9:30

Tim Donohue [9:03 AM]

On the DSpace 7 and Entities side of things, no major updates to share here. Both teams are working hard towards a Preview Release. The Preview Release timeline is still being finalized (it was late Jan, but now looks more like sometime in Feb...will let you know when I know more)
Any specific questions though on DSpace 7 or Entities stuff? That's really all I had to share today...there's plenty more details in the notes of these group's meetings though
Not hearing any, so we'll keep things moving along here
I also don't have any updates on DSpace 6.x releases (6.4 in particular). Again, while I fully suspect there will be more 6.x releases, I think they are on hold until we find a Release Coordinator (or two) that is ready to package up some recent bug fixes into a new release.
Any questions/comments on 6.x releases today?
Ok, same old same old so far. These are all just general updates. Let's move along to updates on recent discussions/activities
@mwood: any updates to share on the Upgrading Solr efforts? <https://wiki.duraspace.org/display/DSPACE/Upgrading+Solr+Server+for+DSpace>

Terry Brady [9:09 AM]

For the Solr work, what are your planned steps. I will resume the work I started on the Docker and schema side once I understand how it my synch with your efforts.

Mark Wood [9:09 AM]

Sharding will have to be done differently. Right now the code rummages through a specific directory looking for shards, and after the upgrade those directories may not even be mounted on the DSpace host.
I need to add a checklist to the PR so that others can see what is being thought of and whether anything has been missed.
One item on that list, that I want to do very soon, is to take up @terrywbrady's schema updates.
I need to look through all of the CLI commands that touch Solr to see if we are doing anything else that assumes direct access to Solr's files.

Tim Donohue [9:11 AM]

I think we should also keep in mind that *everything* need not be solved in a single PR. I'd recommend we get the major update steps into the existing PR, and leave some tasks (e.g. even creating new migration tools etc) for followup PR(s).

Mark Wood [9:12 AM]

If anyone knows of such requirements, or wants to help look for them, that would be helpful.
Yes, @tdonohue, that's a good idea. I should get *something* ready to go in and list next steps.

Tim Donohue [9:13 AM]

So, my point is that a "big picture" checklist of tasks is very appropriate, but the first PR may just tackle the first 1-3 items on that checklist, with followup PRs coming later

Terry Brady [9:13 AM]

It would be reasonable to ignore sharding in an initial release.

Tim Donohue [9:14 AM]

I believe we had also discussed (brainstormed?) possibly leaving sharding to Solr tooling (i.e. DSpace won't do it for you anymore). But, I think we need to look into what that'd mean, and point people to the Solr tooling (examples) that would replace what is in DSpace.

Terry Brady [9:15 AM]

With that approach, we would need to bring the shards home to new repo. Again, that would not be essential preview functionality in my mind.

Mark Wood [9:16 AM]

Time Routed Aliases was what I had in mind. In cloud mode, Solr can be configured to just make new shards as needed. There shouldn't be any manual steps beyond setting it up. (edited)

Tim Donohue [9:16 AM]

@terrywbrady: I agree with you. None of the sharding stuff needs to be figured out by a Preview Release. I think the goal for Preview Release would simply be to run DSpace on a new Solr (with new data, no migration tools or anything)

Mark Wood [9:16 AM]

OK, sharding goes toward the end of the list.

Tim Donohue [9:17 AM]

migration tools, sharding & cleanup of CLI tools can all be left to "followup PRs". I think the first PR should simply concentrate on a *fresh install* of DSpace 7 on new Solr

Mark Wood [9:17 AM]

https://lucene.apache.org/solr/guide/7_4/time-routed-aliases.html is what I was looking at, but I'll take it up later.

Making a note: fresh install first.

Tim Donohue [9:19 AM]

:+1: I think it's fine/good to start thinking about how we'd solve these questions now. But, I'd just prioritize finishing up the "fresh_install" first. Still, if ideas or tools come to mind, jot them down on the wiki page or similar

Mark Wood [9:19 AM]

Will do.

Tim Donohue [9:20 AM]

Sounds good then. So, in terms of "fresh install", it seems like we're decently far along...but I know merging in @terrywbrady's schema fixes still needs to happen. Are there other outstanding needs here (or ways others can help)?

Mark Wood [9:21 AM]

Schema: I just need to do it. I'll make that next step.

Terry Brady [9:22 AM]

Give me a shout if you need me to further enhance the schema fields as you integrate that work.

Mark Wood [9:22 AM]

Will do.

Tim Donohue [9:22 AM]

Ok, it sounds like we've got a plan in place. I'd still recommend adding a checklist to the current PR for tasks that need to be done to achieve "fresh_install" (and maybe it's just the schema). At the very least, it gives us something to refer to in these weekly updates

Mark Wood [9:23 AM]

OK. Will work that out and install a list.

Tim Donohue [9:23 AM]

Any last thoughts/questions on the Solr upgrade? It sounds to me like we are wrapping this discussion up

Mark Wood [9:23 AM]
None from me.

Terry Brady [9:23 AM]
im good

Tim Donohue [9:24 AM]
Thanks @mwood for your continued work here. Moving along then
Next up is any updates on Docker + DSpace from @terrywbrady: <https://wiki.duraspace.org/display/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals>

Terry Brady [9:25 AM]
I am still waiting for reviews. @Patrick Trottier started a couple reviews. I encouraged him to note when his review is done and +1 if things look good.
@jcreel256, if you have time or interest, it would be great to get your input on these.
<https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/68>
<https://github.com/DSpace/DSpace/pull/2307>

Tim Donohue [9:25 AM]
Sounds good. Glad to have @Patrick Trottier back and getting involved again. I wish I could help on the review side, but I'm struggling to keep up on DSpace 7 code reviews right now.

Terry Brady [9:27 AM]
For the DSpace-Docker-Images repo, I have been loose on review requirements. I am holding this one up for review approvals since it is a significant change in how we use docker-compose. I think it is a clear improvement, but I want to get validation from others.
I am meeting Friday with 2 DCAT members to do a repository manager overview of Docker. I will use the feedback from that session as input for the webinar that @pbecker and I will lead on Docker.

Tim Donohue [9:28 AM]
nice, sounds great

Terry Brady [9:28 AM]
We are still working on the language to go out in the webinar invite.
That is it for me

Tim Donohue [9:29 AM]
@terrywbrady: Remind me again on the timelines for your Docker + DSpace webinar? I'm not sure others (especially anyone lurking) is aware that it is coming soon(ish). So, it might be something just to mention briefly now (even though I know there's more advertising to come)

Terry Brady [9:31 AM]
Mar 5 is our working date. It will be a DuraSpace webinar.
The audience will be Repo Managers and developers.

Tim Donohue [9:31 AM]
Thanks again...looking forward to it
Ok, so it sounds like that wraps up Docker discussions for today.
Are there other topics / discussions / PRs to highlight that folks would like to bring up today?

Mark Wood [9:32 AM]
Anything from or for @jcreel256 before he has to leave?

Tim Donohue [9:33 AM]
nothing I can think of, unless he has questions :wink:
I do have a few PRs that I've had in progress to mention here...as I'd like feedback from the general Development community on both.
First, I mentioned this on #rest-api, but I've been working on enhancing our DSpace 7 REST Contract with our documented design principles, terminology definitions, better examples, etc
<https://github.com/DSpace/Rest7Contract/pull/48>
I'd like others to give this a read to see if its an improvement / provides more clarification. Even new/fresh eyes would be welcome, as the goal is to better document this for folks who've never used the REST API before
So far it's really just updates to the main README

Mark Wood [9:35 AM]
pull/48 is on my list.

Terry Brady [9:36 AM]
I will review it as well

Tim Donohue [9:38 AM]

Second, I wanted to also remind folks of an effort to (potentially) release the DSpace 7 backend as a single webapp (combining REST API v7, SWORD, SWORDv2, OAI and RDF into one webapp): <https://github.com/DSpace/DSpace/pull/2265>

I've been revisiting this as late as it's gotten more support/mentions from the DSpace 7 team. So, it's seemingly more likely to occur, but I'd still appreciate positive/negative feedback on it before I begin diving deeper

Currently the PR *just* merges SWORD + REST APIv7 -- so, other webapps are still to come, but you'd get a good idea of the big picture here just by looking at how those two are being merged

And, I'll also note, I've recently realized that merging these webapps may have another major benefit -- Integration Testing across all these webapps. The REST API v7 webapp has very good integration testing framework (from Spring Boot), that we could now apply to SWORD, OAI, etc.

Those are the two PRs I wanted to bring up here. Glad to hear questions or feedback though on either or both

Mark Wood [9:41 AM]

I've made a note to look at it.

Terry Brady [9:41 AM]

These seems like a good idea. I see that this PR mentions that the legacy rest app would not be part of the 7 release. Is that definitely happening?

Tim Donohue [9:42 AM]

@terrywbrady: to clarify, the legacy REST app would *still* be part of DSpace 7. It just would *not* be part of the merged webapp. So, you'd have to install the legacy REST app separately (just like in DSpace 6) if you want it.

Mark Wood [9:42 AM]

There would be little point in doing the work to adopt old-REST into the one-webapp patch, since it *is* going away eventually.

Terry Brady [9:43 AM]

That makes sense.

I do not know much about sword. Is there reason to deprecate the v1 since v2 exists?

Tim Donohue [9:44 AM]

In any case, I do welcome you all to take a closer look. I can also try and bring updates to this meeting as I have any to share. I expect the "one webapp" PR will be a work in progress for some time still,

Mark Wood [9:44 AM]

Someone had a need for SWORD v1 on one of the lists just the other day.

Tim Donohue [9:45 AM]

SWORDv1 is still in use by third-party tools. So, SWORD themselves haven't deprecated it, and until tools move away from using SWORDv1, I think we'll need to continue to support it. I admit though that I wish we could just support one version of SWORD :slightly_smiling_face:

Mark Wood [9:46 AM]

Does "SWORD themselves" still exist?

Tim Donohue [9:47 AM]

"kinda". They actually have a SWORDv3 in the works. But, they aren't really an entity (organization), more of like a project

<http://swordapp.org/2018/05/swordv3-at-open-repositories-2018/>

Mark Wood [9:47 AM]

Yeah, my impression is that JISC just gets a bright idea, shops it around for a reference implementation, writes a report, and it's done. Eventually there's another bright idea....

Terry Brady [9:47 AM]

If the option exists to include in the single webapp or install separately, that could implement an opinion on the matter. Again, I do not know enough to know if this is an important distinction to make.

Tim Donohue [9:50 AM]

@terrywbrady: true. I decided to merge both SWORD webapps in simply because they are rather small overall. Neither are as big as OAI is. That said, they are still separate Maven Modules, so at the point we decide to "deprecate"/remove either one, we can extract the Maven Module easily.

You'll find when you review that "one webapp" PR that the code changes are actually rather small. Currently it's around -75 lines (+229 / -316) to merge SWORDv1 into REST API v7.

And much of that initial refactoring is making this webapp merger possible. I expect merging additional webapps to be fewer changes overall.

It likely won't stay in negative changes though! :wink:

Terry Brady [9:52 AM]

Since I have never seen Sword in action, it would be interesting to create a docker compose file that runs a sword client alongside DSpace. Perhaps both a v1 and v2.

Does such a client exist?

Tim Donohue [9:53 AM]

DSpace XMLUI has a built in SWORDv1 client. But, yes, I agree

I'm not aware of many SWORDv2 clients -- there surely are some, but I think that's part of the reason why SWORDv1 is still in wide use

James Creel [9:54 AM]

The use case for the XMLUI SWORDv1 client isn't super strong, as I recall. You could also do export/import, harvesting, other things.

But Vireo uses SWORDv1

Tim Donohue [9:54 AM]

SWORDv2 implementations: <http://swordapp.org/sword-v2/sword-v2-implementations/>

James Creel [9:54 AM]

Just a curl request and a METS structure sitting next door could suffice for demonstration

Tim Donohue [9:55 AM]

Yes, technically you can send content to either SWORDv1 or SWORDv2 via `curl`. We have examples of both in our DSpace Docs

Mark Wood [9:55 AM]

I'm beginning to imagine a bunch of projects standing around, all thinking "SWORDv1 should be deprecated," but none of them wanting to be the first to do so.

Tim Donohue [9:56 AM]

Example `curl` commands for SWORDv1 are in our docs here: <https://wiki.duraspace.org/display/DSDOC5x/SWORDv1+Server#SWORDv1Server-DeposittoSWORDServer>

Mark Wood [9:56 AM]

Meanwhile ResourceSync is hovering in the background, looking for an opening.

Terry Brady [9:56 AM]

I started a github issue. I will copy this notes there. <https://github.com/DSpace-Labs/DSpace-Docker-Images/issues/73>

Tim Donohue [9:56 AM]

Example `curl` commands for SWORDv2 are in our docs here: <https://wiki.duraspace.org/display/DSDOC6x/SWORDv2+Server#SWORDv2Server-DeposittoSWORDv2Server>

Seems like we're wrapping this discussion up, and I see we are nearing top of the hour anyways

Any final comments/questions for today?

Mark Wood [9:58 AM]

I have none.

Tim Donohue [9:58 AM]

OK, not hearing any. Thanks for the discussion today all!

Have a great rest of your week!

Terry Brady [9:58 AM]

Have a good week!

Mark Wood [9:59 AM]

Thanks, all!