DevMtg 2019-01-09

Developers Meeting on Weds, January 9, 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 10 at 15:00 UTC
- DSpace 7 Entities Working Group (2018-19): Next meeting on Tues, Jan 22 at 16:00 UTC

    - Yesterday's meeting notes at 2019-01-08 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. OR2019 Presentation Planning (EXTENSION: proposals now due on Jan 16)
    a. See 2019-01-10 DSpace 7 Working Group Meeting agenda for a list of the proposed DSpace workshops/presentations (concentrating on DSpace 7)
4. Upgrading Solr Server for DSpace (Any status updates?)
    a. PR https://github.com/DSpace/DSpace/pull/2058
5. 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/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
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 Mark H. Wood)
        i. PR 2180 improves reporting. Ready for review.
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

Meeting Notes

Meeting Transcript

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

```
@here: It's DSpace DevMtg time. Agenda at: https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-01-09
As usual, let's start out with a role call of who is here


Mark Wood [9:00 AM]
I am.


Pascal Becker [9:01 AM]
Hello. :wave:


Terry Brady [9:01 AM]
hello


Tim Donohue [9:01 AM]
Welcome all back from the holidays, and happy new year to everyone.  While there's a lot on the agenda, many of
these topics may not have many updates today (cause of the holiday break)
but let's dive in
I don't have any updates to share on DSpace 7.  Development has been ongoing (lots of PRs under review right
now), but the first meeting of that team is *tomorrow* at 15UTC
The DSpace 7 Entities WG met yesterday, notes are at: https://wiki.duraspace.org/display/DSPACE/2019-01-
08+DSpace+7+Entities+WG+Meeting   That team is still actively working to get an Entities PR ready for the
`master` branch.  Getting much closer, and I expect that PR will be ready before end of Jan (and maybe in a few
weeks)
That's it for DSpace 7.   On the DSpace 6.x side, I don't have any specific updates to share their either.
Obviously there's been more PRs coming in for bug fixes, etc.  We'll likely want to start talking/thinking
about 6.4 soonish, but I'm not sure we're ready for that today
I will however note that I'll want to (eventually) find a 6.4 Release Coordinator. I won't be able to take this
role myself (with all the DSpace 7 work going on)
So, please start thinking about whether that's something of interest to you (not just those here now, but
anyone reading these logs later)


Pascal Becker [9:05 AM]
May I ask: regarding DSpace 7 and deletion of eperson, is there a plan or are there some news?


Tim Donohue [9:07 AM]
@pbecker: last we talked about that, all agreed we need this feature.  There were concerns about the
implementation (setting to null). I asked those with concerns to comment on the ticket & bring discussion
there: https://jira.duraspace.org/browse/DS-4036
However, I see that no one has done so yet.


Pascal Becker [9:08 AM]
Then we're on the same page. :wink: (edited)


Tim Donohue [9:09 AM]
Yes, so we'll have to revisit this again with the DSpace 7 team.  Hopefully we'll find time tomorrow, but
tomorrow's meeting will need to concentrate on OR2019 proposals (as those are due next week) and plans for
Preview Release. Hopefully we can squeeze in 4306 though, or remind folks about it
Anyone else have specific questions on DSpace 7 or 6.x updates?  That's all I had to say there, but obviously
there's other topics on the agenda with some overlap


Terry Brady [9:10 AM]
I am glad the Mirage2 build issues are solved.


Tim Donohue [9:11 AM]
```

Next up was a simple reminder that OR2019 Proposals are due next week, *Weds, Jan 16* (this was extended from an initial due date of today). I don't expect it to be extended further, so get your proposals in!
If you are interested in current DSpace 7 proposal plans for OR2019, they are listed in tomorrow's DSpace 7 mtg agenda (near the top): https://wiki.duraspace.org/display/DSPACE/2019-01-10+DSpace+7+Working+Group+Meeting   If anyone is interested in chipping in on those shared proposals, let me know or attend the DSpace 7 mtg tomorrow
Anyone have anything to discuss (or brainstorms for proposals that you'd like feedback on?) regarding OR2019?
no one is typing, so assuming no :wink:

Pascal Becker [9:14 AM]
Just let us mention, that Terry and me handed in the docker workshop :slightly_smiling_face:
I'm working together with Erin to get the presentation done about the relationship between national user groups and their international communities.

Tim Donohue [9:15 AM]
moving along to next topic. I wanted to check in (with @mwood) on the Solr upgrade for DSpace 7. https://wiki.duraspace.org/display/DSPACE/Upgrading+Solr+Server+for+DSpace , PR at https://github.com/DSpace/DSpace/pull/2058
Is there more support/help/brainstorming needed on this effort? Just wanted to see if this is waiting on feedback or anything

Mark Wood [9:17 AM]
I've been wondering whether we ought to try to patch up the core upgrade task in Ant one more time (the LAST time) or just rip it out.

Terry Brady [9:17 AM]
Fyi... there is some schema work in this PR: https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/65/files

Mark Wood [9:17 AM]
Yes, @terrywbrady I need to look that over and see about merging it into what I'm doing. (edited)

Tim Donohue [9:18 AM]
@terrywbrady: yes, thanks for the reminder...it should definitely get pulled into this work :wink:

Terry Brady [9:18 AM]
The schema changes are incomplete, but it does provide an illustration of how to use the api to alter the schema. There is a preliminary mapping for several fields.
Eventually, the Docker image should pull the schema mods from the DSpace github repo.

Tim Donohue [9:19 AM]
@mwood: do you see any benefit/issues with having Ant help with the upgrade (one last time)? I guess I'm curious if we'll hit issues where Ant doesn't have permissions to do the upgrade, or other such oddities (now that the Solr server is separate)
@mwood: I guess I'm saying, I trust your judgment here. I'm not against ripping out the Ant scripts if they are more trouble than they are worth. But, we'd need to replace them with detailed documentation -- so if automating this upgrade is *easier* than detailed docs, that's great too

Mark Wood [9:20 AM]
Supposing that we do provide one last upgrade, it would happen *before* the cores are handed over to the external Solr instance, so we still have control.

Tim Donohue [9:21 AM]
oh, I see. in that case, it seems like it might be nice to help people along... So, keep the Ant scripts to do one last upgrade, then they can copy them / move them elsewhere (as needed) and not have to worry about manually upgrading first

Mark Wood [9:21 AM]
Hm, it's easy to be lazy and say "sites will have to do this on their own in the future, so make a clean break now and let them start learning."
OK, I'll continue working on The Last Upgrade.

Tim Donohue [9:22 AM]
we said the opposite :wink: I see both sides here, I just tend to lean towards "help with transition" than "force them to learn" for DSpace 7 -- just because DSpace 7 has a *lot of changes to learn about*

Terry Brady [9:23 AM]
I just spent my weekend upgrading stats shards to use UUID's as a part of our DSpace 6 upgrade. We will need to determine if that migration will still be runnable before or after the migration to an external solr.

Mark Wood [9:23 AM]
We also need to settle on a minimum required version of Solr. Latest stable, 7.0, somewhere in between? This is not urgent, but we need it by the time the Release Notes are being written.

Tim Donohue [9:24 AM]
@mwood: I'd go with latest stable unless someone can argue why we shouldn't do so.
I.e. let's get them up to current so they don't need to worry about Solr for a bit.

Mark Wood [9:25 AM]
We might want to leave some leeway for sites that are already running Solr 7x, but not the *very* latest, for
other uses.
The current code wants 7.2 because that's what I had installed at the time.

Tim Donohue [9:26 AM]
That seems reasonable. I'm fine with saying 7.x (in general), and not specifically 7.3.x (assuming that's
easily possible). I just don't see a good reason to say 6.x and above.

Mark Wood [9:27 AM]
Right, I wouldn't set the minimum below 7.

Tim Donohue [9:28 AM]
Sounds like agreement then. I'd say look into support 7.x or above. If for some reason that's problematic and
we need to say 7.2.x or above, then that's OK too (but we should document why we require 7.2.x or above, if so)

Mark Wood [9:28 AM]
OK, I'll back things down to 7.0, but I'll be testing on 7.3 or better.

Tim Donohue [9:29 AM]
Any other final feedback you need on this Solr work? Sounds like you have next steps planned out, but just
wanted to check to be sure there's nothing else you need

Mark Wood [9:29 AM]
Nothing comes to mind. I'll ask if something comes up.

Tim Donohue [9:29 AM]
Sounds great, and thanks again for taking the lead here @mwood
Moving along now. Next up is DSpace Docker updates (from @terrywbrady) https://wiki.duraspace.org/display
/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals

Terry Brady [9:30 AM]
I recommend that we focus on the PR's.

Tim Donohue [9:30 AM]
I know @terrywbrady you've done a lot of recent work here...I admit, I haven't had a chance to catch up on it
all yet, but it look interesting

Terry Brady [9:31 AM]
The following PR is a refactor/cleanup of how we use docker compose: https://github.com/DSpace-Labs/DSpace-
Docker-Images/pull/68
The details are summarized here: https://github.com/DSpace-Labs/DSpace-Docker-Images/blob/helper_cmds/docker-
compose-files/dspace-compose/ComposeFiles.md
With this change, we will have a base docker-compose file for DSpace and then variant files for each major
version of DSpace.
It is also possible to invoke the docker file build process through docker compose. This would allow you to
build and run from the same directory.
I need some folks to give this a try and share their reactions.

Tim Donohue [9:34 AM]
I'll say the concept sounds nice to me, and it sounds like it simplifies the configuration (separates out
changes per version) quite nicely!

Terry Brady [9:35 AM]
@pbecker will you be able to give this a test?

Tim Donohue [9:35 AM]
@pbecker: as I know you work with Docker more frequently, is this something you'd have time in teh near future
to look at?
Thanks, I see you volunteered via emoji above :wink:

Pascal Becker [9:35 AM]
I have it on my todo list, but cannot say yet when I'll have the chance to. Hopefully next week.

Terry Brady [9:35 AM]
Thanks from me as well.
I will alos ask @kshepherd if he can test.

Pascal Becker [9:36 AM]
no problem. It is great what Terry is doing in this area and I'm happy to support that as good as I can.

Terry Brady [9:36 AM]
I will see who else is showing up in the #dspace-docker channel and try to find another tester.
Shall I move on to the 2nd PR?

Tim Donohue [9:37 AM]
I'm also very impressed with all the hard work on Docker. It's something I still want to find time to dig into
& *use* myself. I admit though that I've found very little time these days to actually do coding, so I'm
waiting for a time when I actually have free time again :wink:
Go ahead @terrywbrady

Terry Brady [9:38 AM]
This PR speeds up the Docker build process: https://github.com/DSpace/DSpace/pull/2307

Pascal Becker [9:38 AM]
@tdonohue maybe you're lucky and can join the docker workshop at OR19. :wink:
I guess, you will have a workshop yourself in the same time slot unfortunately. (edited)

Terry Brady [9:38 AM]
It assumes that we create a base dependency dockerfile for each version of DSpace. That base image has already
cached maven dependencies so subsequent builds run faster.

Tim Donohue [9:38 AM]
@pbecker: if I'm available at that timeslot, I will. But, yes, I might end up with a conflict
@terrywbrady: nice idea!

Terry Brady [9:39 AM]
The build also deletes unneeded build artifacts in order to keep the docker images smaller. The PR has some
notes about the improvements.
We already have a 3 part build (maven, ant, copy to tomcat) in order to reduce the size of the final image.
This PR reduces the sizes of those interim images.

Tim Donohue [9:40 AM]
Is there any dependencies between this PR #2307 and the previous work? Or are these standalone enough to be
tested/analyzed/merged separately?

Terry Brady [9:40 AM]
They can be tested separately.
The docker-compose build option in the first PR will be more attractive once this PR is in place.

Tim Donohue [9:41 AM]
ok, so ideally this PR #2307 is merged *first*, but not required (edited)

Terry Brady [9:42 AM]
Exactly.
I will move to the 3rd PR. This one is more experimental.

Tim Donohue [9:42 AM]
Is there a reason this PR #2307 still has a WIP flag?
(work in progress)

Terry Brady [9:43 AM]
Basically, I wanted to share the concept here first. If you all like the general idea, I say we should remove
the flag.

Tim Donohue [9:43 AM]
I'll remove it. The idea seems sound. Just needs reviews/testing

Terry Brady [9:43 AM]
The 3rd PR would allow us to add a docker build to our travis build. This is dependent on 2307 (for speed
purposes). https://github.com/DSpace/DSpace/pull/2308 (edited)
This one might be worth holding until we have more developers actively relying on Docker. The beauty of this
would be that a testable version of every PR would be available.

Tim Donohue [9:46 AM]
This seems like a nice idea, but I agree with your analysis. Are there any limits set by DockerHub that would
cause issues here? For example, any limits to the number of images per day we can push? (edited)

Pascal Becker [9:47 AM]
Just to make this clear: this would speed up PR testing extremely.
You would not have to build PRs locally as long as Travis is green and no rebase necessary. (edited)
You could pull the build image and test it.

Terry Brady [9:47 AM]
I am not sure about that. Given the size of the images, I could imagine that they have some limit. I will ask
their support team.

Pascal Becker [9:48 AM]
@terrywbrady please correct me, if I didn't got the idea correctly.

Tim Donohue [9:48 AM]
Could we get documentation (even basic notes) around *how* PR testing would be done once this PR is in place?
I'd like to see the whole picture here

Terry Brady [9:48 AM]
@pbecker you have the idea exactly.

Pascal Becker [9:49 AM]
$ docker run dspace/pr/123 ; lynx http://localhost:8080/jspui

Tim Donohue [9:49 AM]
Please add that info to the PR description (is what I'm asking) :wink: As I think that could help gain support
from others who use Docker (or could be convinced to use Docker)

Terry Brady [9:49 AM]
docker-compose -e DSPACE_VER=prXXX -p d6 -f docker-compose.yml -f d6.override.yml up -d

Tim Donohue [9:50 AM]
Currently the PR is slightly vague (at least to me)...around the PR testing benefits

Pascal Becker [9:50 AM]
Travis builds the docker image with the pr in place. Someone who want to test it tells docker to run it. Docker
downloads the previously build image and runs it. The image contains a complete installation of DSpace with the
PR in place.
You would not have to run maven locally or wait for maven to finish. Maybe you need to tweak the configuration
accordingly to your tests. (edited)

Terry Brady [9:50 AM]
That would download the image and start a server with the right code. The "-p d6" option would indicate that
docker should re-user volumes created in prior DSpace 6 testing.
Once we figure out how to run all of this in the cloud, we could also test there...

Pascal Becker [9:51 AM]
And if we host the AIPs online we can automate to include those as well.

Tim Donohue [9:52 AM]
To clarify, I do understand the benefits now. I'm just asking that we *document* these benefits in the PR (or
JIRA ticket), as the PR just says "Add Docker push/pull to Travis build" and talks about requirements to do
that. It doesn't note that this makes future PR testing simple via Docker, or how that'd work :wink:

Terry Brady [9:52 AM]
I do want to do a quick update on AIPs.
Good idea @tdonohue. I will add those notes.
Ready to talk about AIPs?

Tim Donohue [9:53 AM]
Sure, we can move along to talking AIPs now. I think this PR #2307 sounds great, just needs clarified docs
(which you will add)

Pascal Becker [9:53 AM]
@terrywbrady will bring DSpace to the point where @tdonohue will get rid of vagrant as Docker is prepared so
well. :wink:
:smile:

Terry Brady [9:54 AM]
Oh, one last benefit of 2307 that I forgot to mention.... it will auto-load AIP files are startup if you set an
environment variable.
We have a repo with 2 very small sets of AIP files. https://github.com/DSpace-Labs/AIP-Files

@pbecker / The Library Code has offered to host larger AIP sets for the project.
I mentioned this effort in DCAT yesterday. I was asking for help with AIP file assembly. I did not yet find
any volunteers.
Later this month, I am going to present DSpace Docker to a couple DCAT members (from Georgetown and U Arizona)
to see if they see the vision from a repository manager standpoint.
My hope is to sell the concept to repo managers. I think if they understand what is possible, we might be able
to find some help with AIP curation.

Pascal Becker [9:58 AM]
I have to run in two minutes sharp.
(next meeting directly following this one)

Terry Brady [9:58 AM]
If you all can think of other rep managers who might like to see this, I would be glad to do another
presentation.

Tim Donohue [9:59 AM]
I wonder if doing a recorded presentation (or webinar) might be useful at some point. The OR2019 workshop also
might be an opportunity to find more AIP help

Pascal Becker [9:59 AM]
@terrywbrady may be it's time for a DuraSpace hot topic webcast?

Terry Brady [10:00 AM]
I'm here if there is interest.

Pascal Becker [10:00 AM]
@terrywbrady shall I reach out to Kristi from DuraSpace?

Tim Donohue [10:00 AM]
Yes, we do have occasional "hot topic" webinars hosted from DuraSpace. I don't know what's lined up this year
yet, but there's always a willingness to schedule/setup webinars (and Kristi is the best contact)

Terry Brady [10:00 AM]
That sounds good to me.

Pascal Becker [10:01 AM]
will do
'bye!

Terry Brady [10:01 AM]
Thanks for the time today!

Tim Donohue [10:01 AM]
thanks @pbecker and @terrywbrady! SOunds like a good idea
Ok, any final thoughts / comments for today? We are at the top of the hour, and it sounds (to me) like we've
essentially wrapped up Docker discussions.
I know we still have ongoing brainstorms/ideas in our agenda, and always the list of PRs to review. But, I'll
leave those for another time, and let you get on with your days
Not hearing anything else, so let's wrap up for today. If other topics/ideas come up, feel free to bring them
to #dev (for ad hoc chat).   Thanks all for the great discussion today! And, don't forget to finish up those
OR2019 presentation/workshop/poster proposals!