DevMtg 2019-02-13

Developers Meeting on Weds, February 13, 2019

Today's Meeting Times

ⓘ

- DSpace Developers Meeting / Backlog Hour: 20:00 UTC in #duraspace IRC or #dev-mtg Slack channel (these two channels sync all conversations)

The meeting on Weds, Feb 20 at 15UTC has been **CANCELLED** (see notes below). Our next meeting will be on Weds, Feb 27 at 20 UTC.

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- 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. Meeting on Weds, Feb 20 (Next Week)

    a. Tim will be unable to attend (conflict with DSpace Steering Meeting)
    b. Should we cancel? Or will someone else volunteer to chair the meeting?
4. 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
5. 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)
6. DSpace Docker and Cloud Deployment Goals (old) (Terrence W Brady)
    a. Build optimization PR reviews:
        i. https://github.com/DSpace/DSpace/pull/2344
        ii. https://github.com/DSpace/DSpace/pull/2345
        iii. https://github.com/DSpace/DSpace/pull/2346
    b. Add Docker build/push to Travis
        i. This make sense to consider after 2307 is merged
        ii. https://github.com/DSpace/DSpace/pull/2308
7. 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.
8. 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
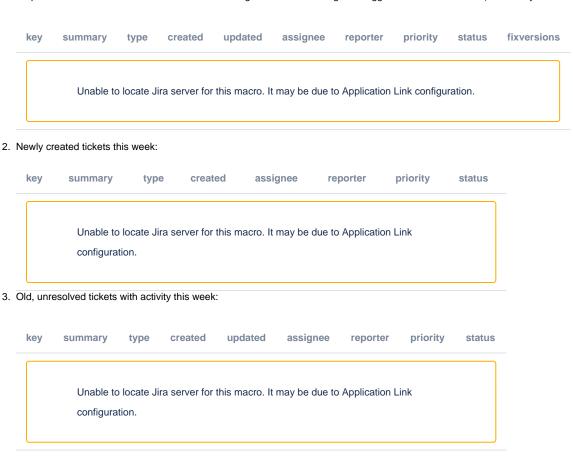
Tabled Topics

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.

Meeting Notes

Meeting Transcript

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

```
Tim Donohue [2:00 PM]
@here: It's time for our DSpace DevMtg.  Agenda is at https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-02-
13
Let's do a quick roll call to see who is here today

Mark Wood [2:00 PM]
I am.

James Creel [2:00 PM]
Howdy

Kim Shepherd [2:00 PM]
hello

Terry Brady [2:01 PM]
hello

Tim Donohue [2:01 PM]
Hi all, welcome.  As I think most of you have been in recent meetings (though, welcome Kim!), the agenda should
look very familiar :wink:
I'll quickly run through quick updates and we'll move along to more specific updates on in-progress PRs, etc.
On the DSpace 7 side, I don't have any special news to report here. Things are highly active, and likely the
meeting notes / meeting itself is the best place to see what's going on with current development
Current goals are still aiming for a 7.0 Preview Release "as soon as possible", and that timeline is looking
more and more like "early March".  But, I'll keep you posted
Any quick questions/comments on DSpace 7 efforts?  Glad to talk more about anything here, as needed
Ok, no one is typing here, so we'll move right along...
DSpace 6.x (6.4) updates.  There's none that I'm aware of.  But, anyone else here have updates/comments
/questions to share?
I'll again assume silence means "no questions" (but definitely interrupt/stop me anytime some come up) :wink:
Moving along into actual topics.  I wanted to note that I'm unavailable for this meeting next week.  On rare
(every few months) occasions the DSpace Steering Group meeting conflicts with this one.
Next week is one of those weeks.  So, should we cancel or would someone else like to take the lead on this
meeting?  (next week it's on Feb 20 at 15UTC)

Terry Brady [2:08 PM]
I may need to miss that one as well.

Kim Shepherd [2:09 PM]
i'll be asleep :wink:

Tim Donohue [2:09 PM]
Ok, since this meeting tends to be sparsely attended these days, let's cancel then.  3 of 5 people here today
are out, so chances are we'd have very few people next week :wink:

Mark Wood [2:09 PM]
OK
```

Terry Brady [2:10 PM]
Fyi, I will be out of the office on the 27, so I will miss that one too.

Tim Donohue [2:11 PM]
Ok, moving right along then. Let's get into PR updates. Solr upgrade is up first (and I know there's been a
lot of activity here this week): https://github.com/DSpace/DSpace/pull/2058
@mwood would you like to start? I know @terrywbrady will likely want to chip in as well

Mark Wood [2:12 PM]
The schemas have been slimmed down as much as I dare, and updated to the latest field implementations since our
old ones were dead and the only other living implementations are deprecated.
'ant fresh_install' seems to be working, as far as it's been tested.
@terrywbrady has been helping me find problems and kick them out of the code.
Ready to talk about the Docker side?

Terry Brady [2:14 PM]
I added a Dockerfile to build a solr container with our 4 DSpace schemas.

Tim Donohue [2:14 PM]
Nice work, both of you! Glad to see this moving more quickly this week

Terry Brady [2:14 PM]
it is deployed as dspace/dspace-solr
The following file allows you to test in Docker with the external solr. https://github.com/DSpace-Labs/DSpace-
Docker-Images/blob/484106a1f6d14e6e37118f7c52faa4f3f6163de7/docker-compose-files/dspace-compose/d7solr.override.
yml
That is part of the following PR: https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/79

Tim Donohue [2:16 PM]
So, this all sounds like good progress. What are the next steps here? Are there more tasks that you both have
on your plates / in your heads? Do we need to search out more reviewers?

Terry Brady [2:16 PM]
I plan to merge the dspace-docker-images pr sooner rather than later.

Mark Wood [2:17 PM]
I'm wondering, at what point to stop and say "we're ready for new installations" and move on to another PR for
supporting upgrades.

Terry Brady [2:17 PM]
I am curious to know some good tests that can/should be executed from Angular or from spring-rest to verify
that Solr is doing what we need it to do.
(@mwood I have not yet tested your latest commits)

Tim Donohue [2:18 PM]
@mwood: yes, that's sorta my question to. I want to know when this PR is considered "finished" enough to put
into the "Needs Review" status (for DSpace 7, etc) and find additional reviewers to help test it out, etc.

Mark Wood [2:18 PM]
I've spent a little time today cleaning comments and commented-outs out of the solrconfig.xml files. I think
there are a number of RequestHandler types that we don't use, and I may try to weed them out separately.
Probably more urgent: I should look over the command-line tools that touch Solr directly and see if they need
work.

Tim Donohue [2:19 PM]
@terrywbrady: Both spring-rest and Angular have built-in integration tests that should give a quick sanity
check. They are not 100% coverage, but they aren't shabby either. So, I'd lean on those a bit first.

Terry Brady [2:20 PM]
I was thinking of manually executed functionality to confirm that solr documents are created and retrieved as
the system needs.

Tim Donohue [2:20 PM]
(And as this PR is built off master, the integration tests for Spring Rest are already passing)

Terry Brady [2:20 PM]
Perhaps those issues would be easier to find as the dspace 7 code base evolves.
I presume some install notes will be needed for developers who are not using Docker to test.

Tim Donohue [2:22 PM]
Possibly, I guess Statistics index is an area that we don't have many (if any) integration tests now. Same

with authority index. The Discovery related stuff has a ton of integration tests though that should be verifying results "look right" when interacting with that index

Terry Brady [2:22 PM]
Unless that is already in place and I missed it.

Mark Wood [2:22 PM]
There are some notes at the top of the PR.

Tim Donohue [2:23 PM]
@kshepherd: since you are here today & have done some past work with Solr, I wonder if you'd also like to give this PR a look/test (at some point) to provide feedback? It'd be a good contribution to DSpace 7 (that only needs knowledge of DSpace 6-ish stuff)

Kim Shepherd [2:24 PM]
yep for sure, i'll test it out this week

Tim Donohue [2:24 PM]
@kshepherd: Thanks! Much appreciated

Mark Wood [2:24 PM]
I'm starting to think we should get this PR ready for final review very soon, give the DSpace 7 crew a heads-up, and try to get it merged, so that people are exercising the code regularly.

Terry Brady [2:25 PM]
For the dspace 7 code base, the docker local.cfg should set solr.server=http://dspacesolr:8983/solr

Mark Wood [2:25 PM]
Anyone working with master will have to pause after the merge to work out how to set up Solr separately. It shouldn't be a big deal.

Tim Donohue [2:26 PM]
@mwood: I'm fine with that. If you feel the PR is in a "good enough" to merge/review stage, we can start that process immediately -- and @terrywbrady and @kshepherd can act as initial testers (and I'll ask for others interested in DSpace 7 mtg tomorrow)
Does that sound like the proper next step? Or is there anything "outstanding" before we bring this to the DSpace 7 team?

Mark Wood [2:29 PM]
I don't think that there are any *necessary* changes left to do in solrconfig.xml; they're just cluttered and messy. I should finish up the mechanical stuff (removing nonfunctional elements) and look at the command line tools in the next couple of days. That should give testers and reviewers time to look over what is done.

Tim Donohue [2:30 PM]
Ok, I'll update the DSpace 7 agenda tomorrow to note this is "ready for review", and see if we have time to even discuss what that means (i.e. once merged, everyone needs to install Solr, etc). That gives that team the option to provide feedback, etc

Mark Wood [2:30 PM]
Good.

Terry Brady [2:31 PM]
fyi, I added a small commit to the PR.

Mark Wood [2:31 PM]
Looks okay to me.

Tim Donohue [2:32 PM]
Ok, sounds like this topic is wrapping up? Any last questions/comments on the Solr Upgrade?

Mark Wood [2:32 PM]
None here.

Terry Brady [2:32 PM]
none

Tim Donohue [2:33 PM]
Ok, moving right along. Next up is my effort to create a "One Webapp Backend" : https://github.com/DSpace/DSpace/pull/2265
I don't have any specific updates here but to note that I'm still working on getting all the webapps merged. Currently it's just RESTv7, SWORDv1 and SWORDv2.... OAI is in progress now.

Mark Wood [2:33 PM]
Drat, I still have a note from last week to inspect that.

Tim Donohue [2:34 PM]
But, it is in a reviewable stage....so, feedback is welcome at any time. I'd really *really* like more "fresh eyes" on this. The DSpace 7 team seems excited, but I want to ensure Committers / this group is too

Terry Brady [2:34 PM]
As the system evolves (DSpace 8 and later), will services like OAI eventually become clients of Spring Rest?

Mark Wood [2:35 PM]
I'm trying to imagine why, as they will have direct access to the business logic.

Tim Donohue [2:36 PM]
@terrywbrady: I think that's yet to be decided...if all these tools should use the REST endpoint for their data.

Terry Brady [2:36 PM]
OK. good to know that is still undecided.
It will be awesome if this improves tomcat startup time

Tim Donohue [2:37 PM]
Regarding this "One Webapp" idea, I'll also mention that in #dev channel this week, @Patrick Trottier came up with a suggestion to name this "single" webapp "DSpace Web Services" (For background, it no longer will be named "spring-rest", as it'll be much more than a REST API)

Mark Wood [2:38 PM]
There is probably a good deal we could do to improve Tomcat startup time by telling it to not scan for various things that we know it won't find -- if Spring Boot hasn't hidden all those controls away somewhere.

Tim Donohue [2:38 PM]
@terrywbrady: it definitely will improve Tomcat startup. It's all in one webapp instead of starting up 4-5 webapps (each with it's own separate Java classes/beans to load, etc)

Terry Brady [2:39 PM]
And I imagine that the final built artifact is much smaller too!

Tim Donohue [2:39 PM]
It also speeds up the Maven build process... less webapps to build & less WAR overlays to process

Kim Shepherd [2:40 PM]
sounds great

Tim Donohue [2:40 PM]
In any case, new/fresh eyes are welcome here. If you look at the PR now, you'll see that SWORDv1, SWORDv2 are merged into Spring-REST...and that I also built out SWORD Integration Tests (basic ones) to prove it works
This exact same concept will be implemented for OAI & RDF (although integration tests might be harder, as RDF needs a triplestore).

Terry Brady [2:41 PM]
thanks for doing this. it sounds like a great improvement.

Tim Donohue [2:42 PM]
No worries, it's something I've had in the back of my mind for a while. And, honestly, Spring Boot makes this a ton easier to achieve (with very little new/refactored code). So, the opportunity is right
That's all I had on this topic though. Any final thoughts/comments/questions?

Terry Brady [2:42 PM]
Could the triplestore be mocked up with an in memory RDF graph for testing purposes?
I don't really have a good sense of what is possible with our mock classes.

Tim Donohue [2:44 PM]
@terrywbrady: possibly, though I haven't looked into it yet. To be fair, I'm not building detailed integration tests for any of these webapps -- those will come later (in a followup PR). The ITs I'm building are simply "does this endpoint respond without throwing a 404? Good!"
But the ITs I'm building could be easily extended / enhanced... I'm just worried that building full ITs will make this PR *massive*. I just want the basic IT infrastructure here...and I can help build more ITs in a future PR
(There are a few other basic ITs I've already built out...like checking that both SWORD endpoints return a valid ServiceDocument. But, for example, I'm not doing a full SWORD deposit via ITs yet)
In any case, that gives you more of a picture of my goals here. Feedback & questions in the PR or on Slack are

more than welcome
Any final thoughts/comments/questions?
Not seeing any typing, so we'll move along
Next up is the ongoing updates on DSpace + Docker: https://wiki.duraspace.org/display/~terrywbrady
/DSpace+Docker+and+Cloud+Deployment+Goals   Anything to update us on this week, @terrywbrady?

Terry Brady [2:48 PM]
The build optimization was merged for 6x last week. The following ports are at +1.

https://github.com/DSpace/DSpace/pull/2344
https://github.com/DSpace/DSpace/pull/2345
https://github.com/DSpace/DSpace/pull/2346
@Patrick Trottier did a nice job of testing these. I am looking for a second +2 for these.
We could talk about automated build of PR's as docker images, but I think that might make more sense after
docker adoption grows.
This PR (on the docker image repo) provides support for externalized solr. it also places every solr repo data
directory into its own docker volume. That way the config is always derived from the image. https://github.com
/DSpace-Labs/DSpace-Docker-Images/pull/79

Tim Donohue [2:52 PM]
Regarding the build optimization PRs...anyone here willing/able to give these a second look? It looks like
these are simply ports of an already merged optimization PR: https://github.com/DSpace/DSpace/pull/2307
(2307 looks to also have been tested by @pbecker, so maybe he'd be available to chip in here at some point)

Terry Brady [2:52 PM]
Also, one quick pitch for the webinar that @pbecker and I will lead on Docker for Repository Managers:
https://duraspace.org/webinar-registration-open-dspace-docker-for-repository-managers-running-any-version-of-
dspace-from-your-desktop/
Duraspace.org
Webinar Registration Open: "DSpace Docker for Repository Managers: Running Any Version of DSpace from your
Desktop" - Duraspace.org
DuraSpace presents a Community Webinar, DSpace Docker for Repository Managers: Running Any Version of DSpace
from your Desktop On Tuesday, March 5, 2019 at 11:00 AM ET (convert to your timezone), join Terry Brady,
Georgetown University Library and Pascal Becker, The Library Code, when they present, "DSpace Docker for
Repository Managers: Running Any Version of DSpace... Read more »
Today at 9:13 AM

DSpaceSlackBot (IRC) APP [2:53 PM]
ogres has quit the IRC channel

Terry Brady [2:53 PM]
Reviews on those PR's would be appreciated! That is all I have.

Tim Donohue [2:54 PM]
Thanks for the updates @terrywbrady and for all the continued hard work on DSpace + Docker!

Kim Shepherd [2:54 PM]
i'll try test out the docker PRs, might ask some questions in #dspace-docker if i get lost, i'm still out of
the dspace 7 loop for the most part

Terry Brady [2:55 PM]
Thanks @kshepherd!

Tim Donohue [2:55 PM]
@kshepherd: yes thanks! FWIW, the Docker stuff isn't all DSpace 7 anyhow... there's docker images for DSpace
4, 5 and 6 too
So, it's a great place to generally chip in... and maybe pick up a few DSpace 7 things along the way :wink:
In any case, thanks again for the updates all. It looks like we're down to <4 minutes
Any final thoughts/topics/updates to share today (from anyone)?

Terry Brady [2:57 PM]
I'll be in and out of the office for much of the rest of the month, so I might be slower to respond than ususal.

Tim Donohue [2:57 PM]
Thanks for the note, @terrywbrady

Mark Wood [2:57 PM]
Noted.
I'll check your "presence light" on Slack before getting impatient. :slightly_smiling_face:

Tim Donohue [2:58 PM]

Ok, not hearing any other topics or updates. So, we'll close up the meeting for today. Reminder, next week's meeting (Feb 20) is *cancelled*. So, the next meeting will be Weds, Feb 27 at 20UTC. (edited)
Have a good rest of the week all! Thanks again

Kim Shepherd [2:59 PM]
cheers all

Mark Wood [2:59 PM]
Thanks, all!

Terry Brady [2:59 PM]
Have a good couple of weeks!