DevMtg 2019-05-22

Developers Meeting on Weds, May 22, 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)
- **Tim is out-of-the-office from May 24-29, 2019.**

This meeting is cancelled next week (Weds, May 29). Tim is out of the office, and others are busy with OR2019 preparations.

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- DSpace 7 Working Group (2016-2023): Next meeting is Thurs, May 23 at 14:00 UTC. Agenda: 2019-05-23 DSpace 7 Working Group Meeting
- DSpace 7 Entities Working Group (2018-19): Next meeting is Thurs, May 23 at 15:00 UTC. Agenda: 2019-05-23 DSpace 7 Entities WG Meeting
- DSpace Developer Show and Tell Meetings: On hold until interesting topics arise.
- 2019 DSpace North American User Group Meeting (Sept 23-24) Call for proposals out!

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))

   a. DSpace 7 Preview Release is out the door!
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 (
   a. **Auto-reindexing in Solr**:   Unable to locate Jira server for this macro. It may be due to Application Link configuration.

      i. Should this only h   need to reindex?
   b.   e authority core.   Unable to locate Jira server for this macro. It may be due to Application

      Link configuration.   Or should we use s
4. DSpace l   p (Tim Donohue)
   a. **First phase was merged**: https://github.com/DSpace/DSpace/pull/2265
   b. A follow-up PR will rename the "dspace-spring-rest" module to "dspace-server", and update all URL configurations (e.g. "dspace.server. url" will replace "dspace.url", "dspace.restUrl", "dspace.baseUrl", etc)
5. DSpace Docker and Cloud Deployment Goals (old) (Terrence W Brady)
   a. Update sequences on initialization

      i. https://github.com/DSpace/DSpace/pull/2362 - update sequences port
      ii. https://github.com/DSpace/DSpace/pull/2361 - update sequences port
   b. DSpace Launcher Dashboard - Deploy a PR on AWS for Testing
      i. There is a 2 minute video that illustrates this proposal.
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
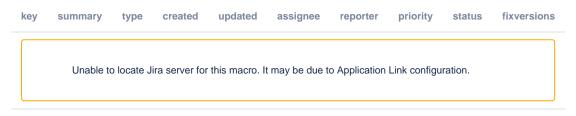
Tabled Topics

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. Brainstorms / ideas
   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)
2. 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.
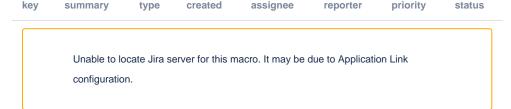
b. In DSpace 6, Hibernate manages the DB connection pool. Each **thread** grabs a Connection from the pool. This means two Context objects could use the same Connection (if they are in the same thread). In other words, code can no longer assume each `new Context()` is treated as a new database transaction.
    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.
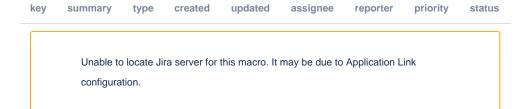
Ticket Summaries

1. Help us test / code review! These are tickets needing code review/testing and flagged for a future release (ordered by release & priority)
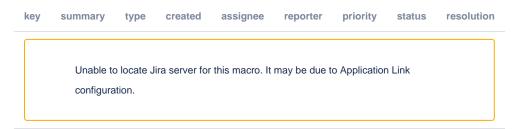
key	summary	type	created	updated	assignee	reporter	priority	status	fixversions

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 CDT)

```
Tim Donohue [3:01 PM]
@here: It's DSpace DevMtg time (a minute late).  Our agenda is at https://wiki.duraspace.org/display/DSPACE
/DevMtg+2019-05-22
Let's do a quick roll call to see who is able to join today

Mark Wood [3:01 PM]
I'm here.

James Creel [3:01 PM]
In a session right now at Texas Conference on Digital Libraries :slightly_smiling_face: (edited)

Terry Brady [3:02 PM]
here

Tim Donohue [3:03 PM]
Ok, let's go ahead and get started
First off, I wanted to add a note in here that I will *not* be able to attend this meeting next week (Weds, May
29 at 15UTC).  I'm taking a short vacation (out of office May 24-29) before the rush to OR2019 in early June
So if there's anyone interested in chairing this meeting next week, let me know (sometime today or tomorrow).
Otherwise, we can always cancel.
(feel free to think on that while we move along to other topics...I'll loop back at the end to see whether we
want to cancel or not)
Moving right along... DSPACE 7 PREVIEW RELEASE is out the door!  https://wiki.duraspace.org/display/DSPACE
/DSpace+7+Preview+Release
Just have to celebrate that milestone, even though I know there's more work to do :tada:
In any case, as folks are playing with the Preview, obviously feel free to send along feedback -- positive
notes, bug reports, frustrating pieces, etc.  It's all good to send along, and it'll help us get things ready
for Beta & final
That's really it for the DSpace 7 update this week.  It's been a big effort, but thanks to the DSpace 7 WG,
DSpace 7 Entities WG and DSpace Marketing WG for helping get this out to the public
In the meantime, the DSpace 7 team is concentrating on OR2019 prep, so there will be code fixes & improvements
in the coming weeks, but they will be oriented towards preparing for OR2019 presentations & workshops.
Any comments / questions on DSpace 7 stuff today?

Mark Wood [3:09 PM]
Includes the One WebApp work, right?
Oh, wait, not in Preview.

Tim Donohue [3:10 PM]
The One Webapp work was merged *just after* the Preview.  It was part of the "prep for OR2019", so it's on the
latest `master`, but not in Preview

Mark Wood [3:10 PM]
Got it.

Tim Donohue [3:12 PM]
Moving right along then.  On the DSpace 6 side, I don't have any updates to note.  Status there is the same as
last time...waiting on volunteers to want to help with packaging up a 6.4
Anyone else with DSpace 6.x updates / comments / questions to share today?

Mark Wood [3:13 PM]
```

Not here.

Tim Donohue [3:13 PM]
Ok, moving along to other updates
@mwood: Any updates to share on Upgrading to Solr Server work? https://wiki.duraspace.org/display/DSPACE
/Upgrading+Solr+Server+for+DSpace

Mark Wood [3:14 PM]
No, I need to get back on that. I've had a stubborn local problem to deal with, plus a week off.

Tim Donohue [3:16 PM]
Ok. I'd be interested to hear any "rough details" on what the upgrade likely looks like, so that I can report
them at OR2019 (in DSpace 7 Workshop).   E.g. Do we feel confident that this is an export & reindex task? If
so, do we need to have a new tool, or can we use an existing one?

Mark Wood [3:16 PM]
I have a note from someone regarding his experience with DSpace + Solr 7, to review soon.

Tim Donohue [3:16 PM]
You don't need to answer those now, but that's the basics of what I'd like to know in the next few weeks :
slightly_smiling_face:

Mark Wood [3:17 PM]
OK, I'll try to have that nailed down soon.

Tim Donohue [3:18 PM]
That'd be nice. If it's not possible to completely nail down, any "best guesses" are also welcome. I just
want to give a status on this, since it's the other possibly "complex" piece of the upgrade (besides the entire
UI, obviously)

Mark Wood [3:18 PM]
OK.

Tim Donohue [3:18 PM]
In any case, moving along for now.
As noted above :point_up: , the first "One Webapp" PR was already merged onto `master`. It's not in the
Preview, but was merged right after release: https://github.com/DSpace/DSpace/pull/2265   This means that all
our separate webapps are now combined in the "dspace-spring-rest" Spring Boot webapp.
There will be a second phase of this work (likely after OR2019) where the "dspace-spring-rest" webapp will be
renamed to "dspace-server" webapp. But, that's mostly just a giant code move -- so, waiting on a slow period
(likely after OR2019) to do that.
That's it for the updates on "One Webapp". The main idea is there on `master` already, but a rename of the
webapp is coming.
Any questions / comments on that?

Terry Brady [3:21 PM]
It will be nice to see that effort completed!

Tim Donohue [3:22 PM]
I agree completely! At this point, it's mostly waiting on my time/availability (as the key driver here). As
noted, I'll be out a bit in the next week, and I'm having to put all my effort to OR2019 prep at this point. :
slightly_smiling_face:
Ok, moving right along to other topics
Next up is Docker + DSpace. Anything you want to update us on @terrywbrady? https://wiki.duraspace.org/display
/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals

Terry Brady [3:24 PM]
Yes, I want to chat about https://terrywbrady.github.io/CldAws230/
terrywbrady.github.io
DSpace Launcher Dashboard
Deploy a DSpace PR in Docker on AWS for Testing
This is an app running in AWS that will spin up a server with DSpace running in Docker. You can select a
specific pull request to initiate.
The app currently allows 2 simultaneous instances to be running. Each instance will be killed after an hour of
uptime.
This is my vision for supporting PR testing by DCAT members and other power users.

Tim Donohue [3:25 PM]
Very cool!

Terry Brady [3:26 PM]

I have implemented this as a class project. If this looks promising, I would like for this to find a home.

Mark Wood [3:26 PM]
A wall-clock hour?

Terry Brady [3:26 PM]
Mark, yes
Tim, consider if DuraSpace might want to make this available.
My days as an active contributor are likely limited. If this looks useful, I would love to find a home for the app.

Tim Donohue [3:27 PM]
What is this little app written in? Are there details on that somewhere? Or is the app code somewhere?

Terry Brady [3:27 PM]
Yes the link above has the code. It is python code deployed as lambdas on AWS.

Tim Donohue [3:28 PM]
Oh, I see it now. overlooked it initially: https://github.com/terrywbrady/CldAws230

Terry Brady [3:28 PM]
I would be glad to do a walk through or a Dev Show and Tell if there is interest.

Tim Donohue [3:29 PM]
Yes, I think that would be awesome. I know I'd like to learn more (after OR2019). I also could promote to other tech folks in DuraSpace/LYRASIS to attend

Terry Brady [3:29 PM]
The branches load really fast. The code currently builds a PR before deployment so those instances take a few minutes to load.
Great. Let's plan to do something in July.
I have about $80 in AWS credits left from my class, so I am glad to use the credits towards demos of the app.

Tim Donohue [3:32 PM]
Sounds great! Hopefully it's something DuraSpace/LYRASIS would be interested in, but need to understand it a bit better first obviously. In any case, seems like an awesome idea & glad to see it works!

Terry Brady [3:32 PM]
Should we schedule after OR or mark a date now?

Tim Donohue [3:34 PM]
I don't have a strong opinion either way, to be honest. We could put something tentative on the calendar & then start talking more at OR (and elsewhere offline). I completely admit, I don't know what my July schedule will look like yet, as we are heading into the "meat" of the merger (actual merger steps gear up in June/July)

Terry Brady [3:35 PM]
I'll propose something. We can adjust the schedule as needed. I am glad you are interested!
I'll do my class presentation in a couple weeks.
I have a tangential issue to mention if there is time.

Tim Donohue [3:36 PM]
Sounds good. Go ahead then on your next topic. We don't have much left on the agenda

Terry Brady [3:36 PM]
@cwilper asked me about automating performance testing. This is related to a ticket he recently raised.
Chris has a set of data that can be used for performance testing DSpace. We might want to figure out a way to do automated performance tests on PR's or on our master branch on a periodic basis.

Tim Donohue [3:38 PM]
I know Chris also has some early notes here: https://wiki.duraspace.org/display/DSPACE/DSpace+7+Performance+Testing

Terry Brady [3:38 PM]
The AWS solution I mentioned above could be useful, or we might want to look at some additional CI tools that could assist us.
I was at a devops meetup and some folks recommend CircleCI which seems to offer some free tier options for open source projects.
This might be a good issue to start tracking in these meetings. Perhaps it could be tackled after the formal DSpace 7 release.

Tim Donohue [3:40 PM]

That's something I'd be interested in seeing happen (automated performance tests on a semi-regular basis -- not sure if it needs to be per PR, but maybe that'd be nice). I admit though I have no experience on how to make that happen...glad others have thoughts though

Terry Brady [3:41 PM]
Neither do I, but it seems much more feasible now that we have an infrastructure for building a test instance with data.

Tim Donohue [3:42 PM]
Yes, agreed. Our Docker setup + our new DSpace 7 test environment in general should combine to make this a bit easier.
In any case, I'm open to the ideas. Prior to DSpace 7 final, I think we definitely need to do some extensive, ongoing performance testing... maybe based on that we can figure out a way to automate it more (even if that ends up being post-DSpace 7.0)

Mark Wood [3:44 PM]
Step 1 is figuring out what is acceptable performance. Maybe step 0 is "what do we mean by performance?"

Terry Brady [3:45 PM]
From a batch processing standpoint: Index X thousand items in under Y min.

Tim Donohue [3:45 PM]
Yes, I think those are both things we can figure out a bit better in the (manual) performance testing of DSpace 7. We may find some parts easier to automate than others, and we'll get a better sense of what "looks good" and what is "too slow" (at least at a UI / REST API perspective)

Mark Wood [3:46 PM]
I would say that performance work starts with "is it good enough that we can release this without embarassment?". After that, "can we make it better enough to be worth the effort of doing so?"

Tim Donohue [3:47 PM]
Yes, I think we're saying the same thing from different perspectives.
Ok, it sounds like this topic is wrapping up. We just have a note to continue to talk Performance testing & start to get our minds around whether any of it can be automated
I realized at the top of this meeting, I forgot to remind folks that the DSpace 7 Entities Working Group is "restarting" tomorrow. It's meeting is immediately after the DSpace 7 WG meeting (in the same place).
Agenda for that Entities WG meeting is at https://wiki.duraspace.org/display/DSPACE/2019-05-23+DSpace+7+Entities+WG+Meeting
Beyond that, I don't think I have any other major topics to bring up for today. Does anyone else here have topics to discuss in the last 9 mins?

Mark Wood [3:51 PM]
You may have noticed that I'm trying to draw attention to a few un-sexy back-end PRs.
https://github.com/DSpace/DSpace/pull/2397 tries to fix our Hibernate cache configuration.

Tim Donohue [3:52 PM]
oh, yes, they are worth mentioning as a reminder to folks here :slightly_smiling_face:

Mark Wood [3:53 PM]
https://github.com/DSpace/DSpace/pull/2116 moves the list of metadata registries loaded at install time out of the code and into the configuration. This may be of interest w.r.t. Entities.

Tim Donohue [3:54 PM]
I'll note that I've made sure both are on the list for the DSpace 7 team too -- with a bolded note about needing more reviews (see #9 and #11 on the list): https://wiki.duraspace.org/display/DSPACE/2019-05-23+DSpace+7+Working+Group+Meeting#id-2019-05-23DSpace7WorkingGroupMeeting-PRsNeedingReview

Mark Wood [3:54 PM]
And I just de-conflicted https://github.com/DSpace/DSpace/pull/1992 again since it addresses a new Jira issue about using configuration variables in email templates.
Thanks for that note.

Tim Donohue [3:54 PM]
Oh, I forgot about that one (#1992). Yes, we should get that on the list too

Mark Wood [3:56 PM]
Folks who have Hibernate performance issues may want to take a look at the caching patch. At least it should make cache tuning efforts actually work rather than being ignored.

Tim Donohue [3:56 PM]
BTW, while we are talking "boring" backend PRs needing review. I have one too: https://github.com/DSpace/DSpace

/pull/2425 This just updates some dependencies to get us off of versions that have known security issues.

Mark Wood [3:56 PM]
That's at +1 from me.

Tim Donohue [3:56 PM]
Yes, #2425 just needs a second reviewer

Mark Wood [3:59 PM]
That's all the patches I'm pushing today.

Tim Donohue [3:59 PM]
Ok, it sounds like we are ready to wrap up today. If anyone here has some extra time for reviewing / testing in the coming days, take a look at the PRs listed above :point_up: None of them are massive, and several are quite easy to review/test quickly.
One final question. Do you all want a meeting on the calendar for next week? Or should I cancel it?
(I know this is deep into OR2019 prep at this point, so if you need the hour to do prep, etc. we can cancel and touch base later)

Mark Wood [4:00 PM]
I can moderate, if we want to meet.

Terry Brady [4:01 PM]
It seems like we covered a lot of ground today. Perhaps we can skip.

Mark Wood [4:01 PM]
I'm okay with that, too.

Terry Brady [4:02 PM]
I'll be traveling 6/5, so OR may be my next time to meet.

Tim Donohue [4:03 PM]
Since this meeting tends to be pretty low attendance anyhow, it sounds like we should cancel :slightly_smiling_face: Honestly, there will not be many updates between here and OR2019....and if you want them, jump into one of the DSpace 7 WG meetings. We can touch base at OR2019
Thanks all! Have a good rest of your week. I'll take next week's meeting off the shared calendar & add a note in today's agenda that we'll cancel

Mark Wood [4:03 PM]
Thanks, all.

Terry Brady [4:04 PM]
Have a good week (or weeks)!