

DevMtg 2019-04-03

Developers Meeting on Weds, April 3, 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, April 4 at 15:00 UTC. Agenda: [2019-04-04 DSpace 7 Working Group Meeting](#)
- [DSpace 7 Entities Working Group \(2018-19\)](#): Next meeting is TBD
 - 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. **NO MEETING on Weds, April 10.** Tim will be at the DuraSpace Summit in St Louis.
2. (Ongoing Topic) [DSpace 7 Status Updates](#) for this week (from [DSpace 7 Working Group \(2016-2023\)](#))
3. (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.
4. Upgrading Handle Server:
 - a. PR: <https://github.com>
5. [Upgrading Solr Server for DSpace](#)
 - a. **Auto-reindexing in Solr:**
 - i. Should this only need to reindex?
 - b. [Authority Core](#)
 - i. [Authority Core](#) or should we use [Solr](#)?
6. [DSpace 7 Spring \(Tim Donohue\)](#)
 - a. PR: <https://github.com/DSpace/DSpace/pull/2265> (PR is finalized & ready for review)
 - 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)
7. [DSpace Docker and Cloud Deployment Goals \(old\) \(Terrence W Brady\)](#)
 - a. PR Build options: <https://github.com/DSpace/DSpace/pull/2385>
 - i. Option 1 - Solve this in the docker build?
 - ii. Option 2 - create feature branch (ie configurable entities) when needed
 - b. Service initialization and docker integration test script
 - i. <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/104>
 - c. Refine Dockefiles for One Webapp
 - i. Keep only spring-rest and rest webapps
 - ii. Optional deployment to ROOT
 - iii. Feedback for one webapp PR (2265)
 - d. 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
 - e. Add Docker build/push to Travis
 - i. We can revisit this after Docker is more widely adopted by DSpace developers. We can decide if travis is the right place to solve this.
 - ii. <https://github.com/DSpace/DSpace/pull/2308>
8. 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)
9. 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 CDT)

Tim Donohue [10:00 AM]

@here: It's time for our general DSpace DevMtg. Agenda is at <https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-04-03>

Let's do a quick roll call to see who is able to join today

Mark Wood [10:00 AM]

Hi.

Terry Brady [10:00 AM]

Partially here... I am listening in to a meeting that is wrapping up.

Bill Tantzen [10:01 AM]

Hey!

Tim Donohue [10:01 AM]

Hi all. Welcome. We'll go ahead and get started :wink:

First off, just a quick announcement. There will be no meeting next Weds (April 10). I'm out of the office next week at the DuraSpace Summit (in St Louis, just after CNRI).

That said, if a few of you want to touch base, you are more than welcome to do so. I just won't be here on the DSpace 7 front. Development is progressing (it always does!) A lot of new code getting merged this week, and we are getting closer and closer to the DSpace 7 Preview Release (current estimate is around April 18).

The major outstanding effort we are waiting on is the final review of Configurable Entities. So, if there's anyone @here still looking to review/test Configurable Entities, we'd appreciate it if you do your initial review **very soon**.

Ideally we'd get this effort merged in the next week to get the Preview Release out the door

Here's the Configurable Entities PRs again:

- (REST Contract) <https://github.com/DSpace/Rest7Contract/pull/57>
- (REST) <https://github.com/DSpace/DSpace/pull/2376>
- (Angular UI) <https://github.com/DSpace/dspace-angular/pull/372>

Terry Brady [10:05 AM]

Where is a good place to look for the list of features to test within the Configurable Entities PR?

Tim Donohue [10:07 AM]

@terrywbrady: the PR descriptions provide that information. And they all link to the (very extensive) Google document that describes everything about Configurable Entities (including screenshots & examples): https://docs.google.com/document/d/1X0XsppZY0tPtbmq7yXwmu7FbMAfLxxOCONbw0_r17jY/edit

Terry Brady [10:07 AM]

Thanks!

Tim Donohue [10:07 AM]

That said, if there's questions **not** answered in the PR descriptions, or something not clear... obviously ask on Slack or in a meeting. Glad to clarify as needed

I think that's it for DSpace 7 updates. Obviously though, more details will be discussed in tomorrow's DSpace 7 meeting.

By the way, if you do plan to review/test Configurable Entities, I'd also appreciate it if you could let me know (or assign yourself to one or more of the PRs as a reviewer). That way I can reach out to see how things are going (and also simply understand who all is planning to review it)

Any questions before we move on?

Mark Wood [10:11 AM]

None.

Tim Donohue [10:11 AM]

Ok, moving along to DSpace 6.x (6.4 to be precise).

As usual, I don't have updates on the code side of things. However, folks from TDL (Texas Digital Library) reached out today to ask if they can provide resources to help out (as they are looking for a few fixes to make their way into 6.x and be released as 6.4)

I, myself, won't have time to devote to coordinating a 6.4 in the near future. But, I thought i'd ask here to see if anyone else has time/interest in coordinating 6.4 (as it sounds like there's resources to pull from to help out). Any thoughts/interest here?

I also plan to email Committers list to ask there (as not everyone is in this meeting). Just haven't gotten to it yet

OK, not hearing anyone jumping up & down here. I'll ask on the Committers list too. Feel free to think on it though, and get in touch if you find time (or your institution is able to free up some of your time to work on it)

Terry Brady [10:15 AM]

I will want to participate, but I cannot offer to coordinate things right now. I need to balance some local projects with DSpace work.

RE 6.4, I was concerned to see <https://jira.duraspace.org/browse/DS-4098>. I had thought most of the performance issues were under control. It looks like there is some good detail in the ticket to help isolate the issue.

Tim Donohue [10:16 AM]

sounds reasonable @terrywbrady.

Terry Brady [10:16 AM]

Are the TDL folks hope to resolve performance, bugs, or something else?

Tim Donohue [10:16 AM]

DS-4098 - yes, it sounds like a problem to me. I honestly haven't verified it personally though.

@terrywbrady: To quote TDL folks: "We have some need of those enhancements and wanted to reach out to see what we can do to help expedite the release. We can devote developer time to the remaining issues (though possibly not the docker issues), but of course our developer can't review and accept PRs, etc. "

I admit, I haven't clarified that statement in any way

Bill Tantzen [10:17 AM]

I can jump in there, as I have been one of the big complainers. I still have performance problems so significant, I cannot upgrade beyond 5.x I will be willing to help out investigating this issue with anybody!

Tim Donohue [10:18 AM]

@bill: Are your issues related to DS-4098? Or something else?

Terry Brady [10:19 AM]

@bill, are your issues on the community-list page or other places? How many collections/items do you have in your repo?

Bill Tantzen [10:20 AM]

Specifically /community-list, but also community and collection pages with a lot of content. But /community-list is the big one, it takes minutes to load.

Tim Donohue [10:20 AM]

One of the first steps here is likely to start **prioritizing** which tickets need to be in 6.4 (vs which are "nice to have"). So, it sounds like there are some bugs / performance issues that are annoying (to at least some folks). I unfortunately don't have a definitive list of those though.

@bill: That sounds like <https://jira.duraspace.org/browse/DS-4098> if you are using the XMLUI. There are some "hints" on that ticket already on things that **might help**. But, I don't think anyone has had time to dig in

deep on that yet (unfortunately)

Bill Tantzzen [10:22 AM]

Right, but I think we have chatted about this before -- the problem exists for me with any theme and with no local enhancements. Plain vanilla.

Tim Donohue [10:23 AM]

@bill: I think that's what DS-4098 is...a plain vanilla issue that is in the *java Aspect layer* of XMLUI. So, it's likely not theme related at all

Bill Tantzzen [10:23 AM]

I'm actually not so concerned about this myself, as I will probably bypass 6.x altogether at this point, but I am concerned that these any similar problems are fixed in 7!

Mark Wood [10:23 AM]

UI problems are fixed in 7 by the existing UIs going away.

Tim Donohue [10:24 AM]

(At least that's my "guess" on DS-4098. I completely admit, I've not had time to verify it. It just sounds like there's a performance bug there)

Terry Brady [10:24 AM]

I have 100 communities + 230 collections and the community-list loads in a reasonable time. It would be interesting to know the threshold in which it gets really bad. Is the issue related to the count of collections or is related to some other properties of those collections.

Mark Wood [10:24 AM]

And Cocoon cache validity stuff is definitely UI code.

Bill Tantzzen [10:25 AM]

Right, so I'm not sure if my problems are related to this particular issue.

Tim Donohue [10:26 AM]

@terrywbrady: I completely agree with you. This issue needs narrowing. It sounds like there *is* an issue, but I'm still not sure how/when it is encountered

Bill Tantzzen [10:26 AM]

For that matter, /jspui is just a bad... I'm convinced it's database or hibernate related but I haven't spent much time with it lately.

Mark Wood [10:28 AM]

OK, that is indeed below the UI layer. But the severity of DS-4098 surprises me. I wonder if we have a very ineffective *Hibernate* cache setup. (edited)

Terry Brady [10:28 AM]

We need to be careful about assuming people will just skip a 6x upgrade.

Bill Tantzzen [10:29 AM]

fwiw, I have 281 communities and 1283 collections.

Terry Brady [10:29 AM]

Wow. That could explain why we are seeing very different results.

Bill Tantzzen [10:29 AM]

sorry, not trying to hijack the meeting with my problems!

Terry Brady [10:30 AM]

Issues like this are really important. They affect the credibility of the platform.

Tim Donohue [10:31 AM]

@terrywbrady: agreed. 6.x performance needs fixing, as I'd actually rather some/most folks upgrade to 6.x first. Jumping from 5.x to 7.x will be possible, but it's a big jump (lots of new API & config changes in 6.x, and then a new REST API and UI in 7.x). Almost all of DSpace (code) changes between 5.x and 7.x

Bill Tantzzen [10:31 AM]

For awhile, Art Lowell was working with me on this, in fact he had a copy of my working db. You might check with him, but likely it has fallen off his radar?

Tim Donohue [10:32 AM]

Art is on 7.x entirely now.

Mark Wood [10:32 AM]

I recall finding that we had multiple EHCACHE config.s that interfered with each other unpredictably. I don't recall whether we fixed that. (edited)

Terry Brady [10:32 AM]

@bill, do your collection objects have bitstreams (logos) attached. I also wonder if that is affecting things.

Tim Donohue [10:33 AM]

Unfortunately 6.x has fallen off all our radars temporarily...7.x development has geared up massively, and we've got few folks looking back at 6.x maintenance currently. 6.x maintenance is definitely coming, but we need to find folks who can "free up" for it.

Bill Tantzzen [10:33 AM]

@terrywbrady very few -- just a handful.

Mark Wood [10:34 AM]

Found it: <https://jira.duraspace.org/browse/DS-3823>

And it is not fixed.

Tim Donohue [10:35 AM]

So, I'm not entirely sure how to move this forward. I don't want this to be 'forgotten', but it's also unclear how to create a ticket here / investigate this.

It sounds like we need to determine if this is at all related to DS-4098 (XMLUI issues). It sounds unlikely.

But, then we need to determine which API layer calls are performing badly

Terry Brady [10:36 AM]

@bill, if you are interested in driving a fix for this, I would be glad to meet and brainstorm.

I will have limited capacity to try to recreate the problem on my end.

Bill Tantzzen [10:37 AM]

OK, I don't want to take up any more time on it here! We can chat out of the meeting maybe? I don't even have a 6.3 instance running anymore so I will have to spin one up...

O

O

I've been focusing on 7.

Tim Donohue [10:38 AM]

Ok, I'll leave this to @terrywbrady and @bill then to move forward / discuss. Hopefully we can eventually get to a ticket to describe the problem, and then look towards a fix. If this *is* a Java API layer problem it might also affect 7.x.

As noted above, I'll also reach out to Committers about coordination of a 6.4 release in the near(ish) future Moving along for now (as we are getting short on time, and I have another meeting at the top of the hour)

Next up: As you may have seen, the team at CNRI has reached out to help us *upgrade our Handle server*:

<https://jira.duraspace.org/browse/DS-4205> and <https://github.com/DSPACE/DSPACE/pull/2394>

Terry Brady [10:41 AM]

It would be good to see if @aroman and @Germán Biozzoli could attend a developer meeting in the future to help keep attention on this issue.

Germán Biozzoli [10:41 AM]

joined #dev-mtg by invitation from Terry Brady.

Tim Donohue [10:41 AM]

This DS-4205 work is for DSpace 7. It'd be good to get some early reviews on, as we all know our embedded Handle server is out of date

However, I will note that the PR#2394 currently doesn't "build" as it's waiting on us to push an updated Handle.jar to Maven Central. I have that on my TODO (haven't gotten to it yet)

(In the long term, CNRI plans to eventually distribute to Maven Central themselves. But, they aren't yet setup to do that, and have given us permission to do so ourselves)

Terry Brady [10:43 AM]

This is a component that would be useful to docker-ize. I will add an issue to our docker repo.

Mark Wood [10:43 AM]

In the long term I would still like to un-embed the Handle resolver.

Terry Brady [10:44 AM]

<https://github.com/DSPACE-Labs/DSPACE-Docker-Images/issues/108>

Tim Donohue [10:44 AM]

I'll also note that @Ian Little (CNRI) who built this PR is also on our Slack. So, if folks have questions as you start to look at this work, feel free to ask here (or in GitHub)
@mwood: yes, I agree. This is just a first step -- at least we'd be on the latest version of Handle Server. Then we could move to un-embed it in 8.0 (or similar)

Ian Little [10:45 AM]

Hi everyone. I usually have Slack on in the background, so feel free to DM me with questions as well.

Tim Donohue [10:47 AM]

So, that's all I had to say about the Handle Server upgrade work. I'll also mention this to the DSpace 7 team, as I'd really like to see this moved along soon (it seems like a small change overall). I'll also let everyone know once I've pushed the latest Handle.jar up to Maven Central (hopefully this week)
Moving right along for now. Next up is updates on the Solr v7 upgrade from @mwood. Any updates this week to share?

Mark Wood [10:49 AM]

I've started work on being able to upgrade an existing DSpace instance to DSpace 7 + Solr 7. Thanks to @bill for early testing that has found problems with date fields.

<https://github.com/DSpace/DSpace/pull/2393>

if anybody wants to track this work-in-progress.

Tim Donohue [10:51 AM]

Thanks for those updates.

Mark Wood [10:52 AM]

The behavior of something changed between SolrDateField and DatePointField. We now have to format a String instead of passing a Date directly to SolrInputDocument.addField. There's discussion on the solr-users list about that now.

Tim Donohue [10:52 AM]

Good to know. Definitely keep us updated (or add details as you learn them into the PR)

Mark Wood [10:52 AM]

Will do.

Tim Donohue [10:53 AM]

For now, I'm going to move along. (I'm short on time myself)

Mark Wood [10:53 AM]

Nothing else on that now.

Tim Donohue [10:53 AM]

On the DSpace 7 "One Webapp" topic, I don't have any updates to share. I've been trying to get back to this work to look into some feedback from @terrywbrady on his Docker testing. Other feedback is still welcome, but I've got no updates <https://github.com/DSpace/DSpace/pull/2265>
So, we can move right along from there

Bill Tantzen [10:54 AM]

@mwood we should continue the discussion on the dev channel later -- I have a few more questions/comments...

Mark Wood [10:54 AM]

I'm in the middle of reading 2265 today.

Tim Donohue [10:54 AM]

@terrywbrady: Any updates you want to add on DSpace Docker deployments, etc.?

Mark Wood [10:54 AM]

@bill OK.

Terry Brady [10:54 AM]

I am going to merge some documentation changes related to memory usage in Docker: <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/107>

I documented a process for running our docker images within an EC2 on AWS.

It would be interesting in DSpace 8 or DSpace 9 to be able to deploy DSpace purely as containers (no command line interaction). I started a ticket to capture discussion on that. <https://jira.duraspace.org/browse/DS-4204>

I'll be happy to re-test the one webapp stuff again. I think I am waiting on a fix for some RDF stuff.

@bill and @mwood are your institutions moving to the cloud providers for deployment?

Mark Wood [10:57 AM]

Not here.

Tim Donohue [10:57 AM]

@terrywbrady: Thanks for the updates. Yes, I haven't gotten back to the "one webapp" stuff in a while now. DSpace 7 reviews & push for Preview release have taken priority. Will ping you once I do get back to it.

Bill Tantzen [10:58 AM]

no, not here...

Terry Brady [10:58 AM]

That is helpful to know. I suspect that DSpace will be one of the more difficult apps to migrate. We are starting that process for some of our test servers.

Tim Donohue [11:00 AM]

I'm personally interested in (someday) replacing demo.dspace.org with a Docker deployment (on AWS). Not sure when that'd happen exactly though, but may be the opportunity to move that way for DSpace 7 demo sites. I'm gonna have to unfortunately close up today's meeting. I have to run now to a DSpace Leadership Group meeting.

Terry Brady [11:01 AM]

Have a great week! @bill give me a shout if you want to brainstorm on the community-list issue.

Tim Donohue [11:01 AM]

As we won't have this meeting next week, I'll talk to you again in 2 weeks (if I don't talk to you in the DSpace 7 meetings)

Thanks all!

Mark Wood [11:01 AM]

Thanks, all.

DSpaceSlackBot (IRC) APP [11:01 AM]

epinky has quit the IRC channel

Bill Tantzen [11:02 AM]

@terrywbrady OK, let's talk. Give me the rest of the day or so to get a v6.3 instance going again with some real data!

Bill Tantzen [1:25 PM]

@terrywbrady Terry, why don't you contact me at tantz001@umn.edu, and send me a client ip address -- my 6.x site is behind a firewall. Good news is that I was wrong -- jspui response time is just fine, so this problem is local to xmlui and not core!