

DevMtg 2019-01-16

Developers Meeting on Weds, January 16, 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)

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- [DSpace 7 Working Group \(2016-2023\)](#): Next meeting is Thurs, Jan 17 at 15:00 UTC
- [DSpace 7 Entities Working Group \(2018-19\)](#): Next meeting on Tues, Jan 22 at 16:00 UTC
 - Last 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. [Upgrading Solr Server for DSpace](#) (Any status updates?)
 - a. PR <https://github.com/DSpace/DSpace/pull/2058>
4. [DSpace Docker and Cloud Deployment Goals \(old\)](#) (Terrence W Brady)
 - a. Simplify invocation by using multiple fragments, auto load content on startup
 - i. <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/68>
 - ii. Summary page: https://github.com/DSpace-Labs/DSpace-Docker-Images/blob/master/helper_cmds/docker-compose-files/dspace-compose/ComposeFiles.md
 - b. Speed up Docker builds
 - i. <https://github.com/DSpace/DSpace/pull/2307>
 - c. Add Docker build/push to Travis
 - i. This make sense to consider after 2307 is merged
 - ii. <https://github.com/DSpace/DSpace/pull/2308>
5. Brainstorms / ideas (Any quick updates to report?)
 - a. (On Hold, pending Steering/Leadership approval) Follow-up on "DSpace Top GitHub Contributors" site (Tim Donohue): <https://tdonohue.github.io/top-contributors/>
 - b. [Bulk Operations Support Enhancements](#) (from Mark H. Wood)
 - c. [Curation System Needs](#) (from Mark H. Wood)
 - i. PR 2180 improves reporting. Ready for review.
6. Tickets, Pull Requests or Email threads/discussions requiring more attention? (Please feel free to add any you wish to discuss under this topic)
 - a. Quick Win PRs: <https://github.com/DSpace/DSpace/pulls?q=is%3Aopen+review%3Aapproved+label%3A%22quick+win%22>

Tabled Topics

These topics are ones we've touched on in the past and likely need to revisit (with other interested parties). If a topic below is of interest to you, say something and we'll promote it to an agenda topic!

1. Management of database connections for DSpace going forward (7.0 and beyond). What behavior is ideal? Also see notes at [DSpace Database Access](#)
 - a. In DSpace 5, each "Context" established a new DB connection. Context then committed or aborted the connection after it was done (based on results of that request). Context could also be shared between methods if a single transaction needed to perform actions across multiple methods.
 - b. In DSpace 6, Hibernate manages the DB connection pool. Each **thread** grabs a Connection from the pool. This means two Context objects could use the same Connection (if they are in the same thread). In other words, code can no longer assume each `new Context()` is treated as a new database transaction.
 - i. Should we be making use of `SessionFactory.openSession()` for READ-ONLY Contexts (or any change of Context state) to ensure we are creating a new Connection (and not simply modifying the state of an existing one)? Currently we always use `SessionFactory.getCurrentSession()` in `HibernateDBConnection`, which doesn't guarantee a new connection: https://github.com/DSpace/DSpace/blob/master/dspace-6_x/dspace-api/src/main/java/org/dspace/core/HibernateDBConnection.java
 - c. Bulk operations, such as loading batches of items or doing mass updates, have another issue: transaction size and lifetime. Operating on 1 000 000 items in a single transaction can cause enormous cache bloat, or even exhaust the heap.

- i. Bulk loading should be broken down by committing a modestly-sized batch and opening a new transaction at frequent intervals. (A consequence of this design is that the operation must leave enough information to restart it without re-adding work already committed, should the operation fail or be prematurely terminated by the user. The SAF importer is a good example.)
- ii. Mass updates need two different transaction lifetimes: a query which generates the list of objects on which to operate, which lasts throughout the update; and the update queries, which should be committed frequently as above. This requires *two* transactions, so that the updates can be committed without ending the long-running query that tells us what to update.

Ticket Summaries

1. Help us test / code review! These are tickets needing code review/testing and flagged for a future release (ordered by release & priority)

key	summary	type	created	updated	assignee	reporter	priority	status	fixversions
Unable to locate Jira server for this macro. It may be due to Application Link configuration.									

2. Newly created tickets this week:

key	summary	type	created	assignee	reporter	priority	status
Unable to locate Jira server for this macro. It may be due to Application Link configuration.							

3. Old, unresolved tickets with activity this week:

key	summary	type	created	updated	assignee	reporter	priority	status
Unable to locate Jira server for this macro. It may be due to Application Link configuration.								

4. Tickets resolved this week:

key	summary	type	created	assignee	reporter	priority	status	resolution
Unable to locate Jira server for this macro. It may be due to Application Link configuration.								

5. Tickets requiring review. This is the JIRA Backlog of "Received" tickets:

key	summary	type	created	updated	assignee	reporter	priority
Unable to locate Jira server for this macro. It may be due to Application Link configuration.							

Meeting Notes

Meeting Transcript

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

Tim Donohue [2:00 PM]

@here: It's DSpace DevMtg time! Agenda is at <https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-01-16>
Let's start off with our usual roll call to see who is joining today (edited)

Kim Shepherd [2:00 PM]

<-- here

Mark Wood [2:00 PM]

Here!

Terry Brady [2:01 PM]

here

Tim Donohue [2:02 PM]

Hi all & welcome :slightly_smiling_face: As noted in #dev, the agenda today is pretty light, mostly a carryover from last week. But, we'll check in on each topic for updates, and we likely should have time at the end for extra discussion

On DSpace 7 side, a few updates to note...

First up, we've got our OR2019 proposals in for DSpace 7. If you want to see what got submitted, they are on this wiki page: <https://wiki.duraspace.org/display/DSPACE/DSpace+7+at+OR2019>
(And don't forget to get those OR2019 submissions in today!)

Mark Wood [2:03 PM]

Mine is done.

Kim Shepherd [2:04 PM]

i got mine in

Terry Brady [2:04 PM]

Mine are in as well

Tim Donohue [2:04 PM]

Next, as discussed in last week's meeting with Steering group (today), the DSpace 7 "Preview" Release is going to be later than expected. It won't make the previously mentioned "late Jan" timeframe

That said, I don't have an updated date yet (I anticipate sometime in Feb). I'm calling a meeting of the main contributors (DuraSpace, Atmire & 4Science) to talk our upcoming schedules -- I need to ensure we're all aligned, as in recent months we hit "slow periods" cause one (or more) key players were unavailable.

So, much more info once I get it.

But, the main story here is that, I don't think we're delayed by much...I just don't know if the delay will be "a few weeks" or more like late-Feb. Plus, I really need to understand the availability of the team in the coming months to ensure our Beta & final stay on schedule

I think that's it for DSpace 7 updates. Any questions or comments?

Oh, and I'll note here... if anyone @here wants to get involved with DSpace 7 (even occasionally), we still really have a need for *code reviewers* or *testers*. Some of our most recent minor delays have revolved around having to wait for reviewers/testers to have time, etc

Kim Shepherd [2:09 PM]

i'm still struggling to find a way to participate with dspace 7 given my time zone, the sprint-based dev process etc

Terry Brady [2:09 PM]

I wish I had more time to offer. My plate seems to keep filling up with work on the current branches.

Kim Shepherd [2:10 PM]

but perhaps i could start off with some out-of-hours code reviewing and testing... if it doesn't have to be done in certain times or with others online

Tim Donohue [2:11 PM]

@kshpherd if you find the time or interest, out-of-hours code reviewing / testing would be a *great* way to start to get involved. I definitely get it that you cannot attend the weekly meetings -- but we don't require

everyone to do so.

That said, it's completely up to you. Just let me know (or do a few ad hoc and see how it goes first)

Kim Shepherd [2:11 PM]

@tdonohue i also can't make sprints, usually.. at least not when anyone else is awake
ok i'll give some testing a go, good way to familiarise myself with the code too

Mark Wood [2:12 PM]

Sprinting ended some time ago, no?

Tim Donohue [2:12 PM]

@kshepherd :We aren't doing "official" sprints right now. We only have one meeting per week, which is more like a weekly checkin on what was achieved (in the last week), what needs help/eyes, and what's coming next

Kim Shepherd [2:13 PM]

oh, my mistake, i thought we were just in an 'in between sprints' gap right now..

Tim Donohue [2:14 PM]

Yes, we had to set aside community sprints for now (to move faster). We'll be picking those back up later this year -- likely after the Beta, as a way to do developer fixes from *testathon*, or similar

So, no Sprints for at least the next few months (Beta is estimated for April)

Any other comments/questions on DSpace 7?

Not hearing any...moving along

I don't have any official news to share on DSpace 6.x (a 6.4 release). Obviously, it'll happen at some point (plenty of good fixes already merged & more coming in daily). But, schedule won't be narrowed down until we have a Release Coordinator identified

So, if anyone *does* want to see it happen more quickly, either volunteer to coordinate or co-coordinate. (I definitely cannot take that role, as I'm swamped with DSpace 7 management)

Kim Shepherd [2:18 PM]

i've been working on a few fixes as some of you know... and plan to try do a bit more "catching up" on PR reviews, tests, merges for 6.4 milestone issues so we don't end up too far behind
but i will admit i found the 6.3 coordination a bit overwhelming sometimes, just due to the sheer number of outstanding issues and moving goalposts, so if this one is a similar size i would recommend at least 2 people sharing coordination duties :wink:

Tim Donohue [2:19 PM]

Sounds good, and thanks @kshepherd. As new PRs come in, I've still been tagging them. So, I'd encourage looking more closely at anything I flagged specifically for 6.4 (as it's likely something I thought might be a good addition), especially any tagged "bug" or "quick win".

I agree, co-ordination is always nice.

Terry Brady [2:20 PM]

I have a non-DSpace project I need to tackle next. Once that is done, I might be able to help. I do not yet know how big it will be.

Now that we are on 6x, it will be much easier to assist.

Tim Donohue [2:21 PM]

Thanks as well, @terrywbrady! Definitely keep us in the loop

Ok, I don't have anything else specific to 6.x to say. Any other comments/questions on 6.x?

No one is typing, so moving along :wink:

Next up is an update on how the Solr upgrade (for 7.x / master) is progressing: <https://wiki.duraspace.org/display/DSPACE/Upgrading+Solr+Server+for+DSpace>

<https://github.com/DSpace/DSpace/pull/2058>

Anything to share this week, @mwood? Or any support you need?

Mark Wood [2:23 PM]

No, I need to push through some local work and then refocus on that.

Kim Shepherd [2:23 PM]

looking at the last update on PR comments, looks like stats shard detection / handling needs some reconsideration with an external solr server?

Mark Wood [2:25 PM]

Yes. I will stare at that a bit. I'm not entirely happy with the way we have bent sharding to do yearly cutoffs. I think that recent Solr may provide a better way. I should take some time to jot down my thoughts for critique.

Terry Brady [2:26 PM]

I linked <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/65> to the wiki page. It has some partial schema work in place.

Tim Donohue [2:26 PM]

Yes, it sounds like a section for discussion could be added to that wiki page. Maybe link in related docs (or articles) on Solr sharding that may be of use to review, etc.

Mark Wood [2:26 PM]

Good. I still have to look at your schema updates.

Terry Brady [2:27 PM]

I modified the easy fields... but it will get some of the work out of the way.

Tim Donohue [2:28 PM]

In terms of timelines, it'd be nice to see an early version of this ready for the "Preview" release. However, obviously, it's possible that we may not have everything answered / documented by then. If the Preview turns out to be hard to achieve though, we can aim it for Beta

Mark Wood [2:28 PM]

OK. Since the Preview has slipped a little, that's a little more time to try to get this in.

Tim Donohue [2:28 PM]

(I just know the earlier we get this ready, the more "polished" it'll end up in the end)
Thanks for the updates, @mwood! I'll pause briefly to see if there's any final comments/questions from anyone
None. So, moving along
Next up is Docker updates from @terrywbrady: <https://wiki.duraspace.org/display/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals>

Terry Brady [2:30 PM]

I am looking for reviews on <https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/68>
and on <https://github.com/DSpace/DSpace/pull/2307>
We had some chat in the Dev channel yesterday on 2307 which led to additional improvements.

Kim Shepherd [2:31 PM]

@terrywbrady yep, i'll definitely be giving this a whirl

Terry Brady [2:32 PM]

@Hrafn Malmquist submitted some PR's for our Docker images. It was great to have another developer working on the files.

Kim Shepherd [2:32 PM]

i was also thinking, subsequent to our chat about web.xml and fresh_install etc., that perhaps there are some docker-specific ant tasks we could work on, to keep some deployment jobs working nicely for docker containers

Terry Brady [2:32 PM]

Interesting thought. That could make sense.

Tim Donohue [2:34 PM]

We could also start looking into (cough cough) getting rid of `ant` :smiling_imp:

Terry Brady [2:34 PM]

DockerHub has a new build system. I think it may have been necessitated by some web hook/integrations that GitHub will be retiring. I added some additional issues to our repo with additional brainstorm.

<https://github.com/DSpace-Labs/DSpace-Docker-Images/issues> (edited)

Tim Donohue [2:34 PM]

And, yes, I mean that seriously -- but not necessarily for DSpace 7 :wink:

Mark Wood [2:34 PM]

Nooo! Ant does procedural stuff so nicely. We won't get Maven to do installation without a struggle.

Terry Brady [2:35 PM]

Those are my updates

Tim Donohue [2:35 PM]

Thanks @terrywbrady.

Mark Wood [2:35 PM]

If I ever get time to return to my installer wizard, I can at least pack Ant inside it where you don't have to see it.

Tim Donohue [2:36 PM]

@mwood: To clarify my comments....I don't want to turn Ant's tasks over to Maven. I'd either pack Ant inside (so it's "invisible") or Java-ify all those tasks.

Mark Wood [2:36 PM]
OK

Tim Donohue [2:37 PM]
But, that's a bigger discussion. Just noting that Ant gets in the way of things now...and Docker is only one recent example. We need to figure out a way to do this stuff better

Kim Shepherd [2:37 PM]
:ant:

Tim Donohue [2:38 PM]
In any case, none of that's planned for DSpace 7. So, we'll table it. I'm just wanting to insert this "early idea" into everyone's heads so we can start thinking about how it might be done better, eventually
Any other questions/comments on Docker stuff?

Kim Shepherd [2:38 PM]
just that i really like it :wink:

Tim Donohue [2:39 PM]
I really like it too (that it's happening)...and I really want to find a few days in the future where I have time to figure out how to *use it myself* :wink:
But, I don't have that sorta free time anymore (not with DSpace 7)
:slightly_smiling_face:

Terry Brady [2:40 PM]
If you have an AWS account, I can share some details on how you can use it on one of their Cloud9 EC2 instances. Ping me if you find the time or need.

Tim Donohue [2:41 PM]
It's more time than anything else on my end (willingness is there). I don't have time in my day to even do code development right now. But, I appreciate the offer.
In any case, moving along now. We're essentially at the end of our Agenda. There are some various "Brainstorms / Ideas" in topic #5, but I'm not sure any have had recent updates
Are there other topics / discussions that anyone here has to bring up?

Terry Brady [2:43 PM]
ArchivesSpace is doing something interesting. <https://archivesspace.atlassian.net/wiki/spaces/ADC/pages/802127927/ASpace+Online+Forum+2019>
They are hosting an online conference to include folks in all time zones. We hope to present something at it.

Kim Shepherd [2:44 PM]
@mwood i still want to help test your curation reporting PR -- are there any test instructions / examples you could share to help me get stuck into it? i'm just not 100% sure i know how to test coverage of all the improvements right now

Mark Wood [2:44 PM]
I'll just mention that I've scribbled some more thoughts at the bottom of Tabled Topics, about bulk operations and rethinking how we use the database.

Tim Donohue [2:44 PM]
@terrywbrady: Cool, that might be something to pass along to the (upcoming) DSpace Marketing Working Group!
(By the way all, there's a DSpace Marketing Working Group that will be established in the next week -- look for a public call on mailing lists as early as possibly Friday)

Mark Wood [2:46 PM]
@kshepherd Near the top there is some background. I'll see if I can come up with more information for testers.

Kim Shepherd [2:46 PM]
one old followup from me: i think it'd be great to have a private DSpace repo on the duraspace account to handle security / sensitive PRs... is that something we could A) ask github if they could give to duraspace for free, or B) raise with steering committee to see if there's money for something like that?
or s/steering committee/duraspace leadership/ i guess? whoever pays for things :slightly_smiling_face:

Tim Donohue [2:48 PM]
@kshepherd: I think we could likely make that happen...*However*, IIRC, I think there's already a private DSpace repo hosted by @pbecker (or there was at one point)
In any case, we can talk more about it and figure out the best place. I'm open to ideas, and I think that'd be pretty low cost if we had to bring to Steering/Leadership

Pascal Becker [2:49 PM]

<https://github.com/the-library-code/Dspace-Security>
You have to be logged-in and allowed to see it (edited)

Kim Shepherd [2:51 PM]

@pbecker cool, if you're ok with me using it then i'll make that my place for sensitive PR reviews :
slightly_smiling_face:

Pascal Becker [2:52 PM]

Go ahead. You should already have access rights. If not, ping me!
I happily add any committer that want to have access.

Kim Shepherd [2:53 PM]

if we wanted to give an issue reporter / expert temporary access, would that be possible without also allowing access to all the other branches/PRs?

Terry Brady [2:53 PM]

Let me know if you need a test reviewer on something.

Pascal Becker [2:55 PM]

No. We can currently give access to the repository or not. As far as I know there is now way to give access to some branches only.
But we can remove old branches.
Have to run!

Tim Donohue [2:55 PM]

FWIW, I don't see any old branches (I think they were already removed). In any case, we can talk more post-mtg, I'm sure
I see we are nearing the top of the hour -- any final thoughts/comments/questions to add?

Kim Shepherd [2:57 PM]

if anyone is going through quick wins in the coming week, a lot of them already have a +1, i've been trying to approve things lately, so hopefully there are quite a few that can be merged

Tim Donohue [2:58 PM]

I'm not sure whether I'll get there this week, but if I find some extra time , I'll give them a look too!
Before we close up, I'll also add a reminder that if anyone has agenda items for future meetings, please feel free to pass them along (or add them to the agenda yourself, if it exists).
Other than that, we may as well close up on time today. I appreciate the discussion today all!

Mark Wood [3:00 PM]

Thanks, all!

Kim Shepherd [3:00 PM]

cheers all!

Terry Brady [3:00 PM]

Have a good week!