DevMtg 2018-10-24

Developers Meeting on Weds, October 24, 2018

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)
    - Please note that all meetings are publicly logged

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- DSpace 7 Working Group (2016-2023): Next meeting on Oct 25 at 14:00 UTC
- DSpace 7 Entities Working Group (2018-19): Next meeting is Nov 6

    - Last meeting (including video recording): 2018-10-23 DSpace 7 Entities WG Meeting
- DSpace Developer Show and Tell Meetings: Next meeting is Oct 30 at 15:00 UTC on Docker Updates

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. DSpace 7.x dependency update PRs
    a. https://github.com/DSpace/DSpace/pull/2235 (Update all webapps to Servlet API Spec v3.1)
    b. https://github.com/DSpace/DSpace/pull/2244 (Apache Commons Configuration v1 v2 and Commons Lang v3)
    c. https://github.com/DSpace/DSpace/pull/2241 (Log4j 1.2 2.x) – Is this ready for review?
    d. https://github.com/DSpace/DSpace/pull/2058 (Solr v7). Tim will be adding some fixes
        i. Dependent on schema changes in https://github.com/DSpace/DSpace/pull/1810
4. DSpace 6x: Merge Legacy and Current Statistics: https://github.com/DSpace/DSpace/pull/1810
5. Docker PRs (Terrence W Brady) : DSpace and Docker#Proposal:SupportMultipleJDKVersions
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. Follow-up on Curation Task Reporting (PR 2180)
    c. Bulk Operations Support Enhancements (from Mark H. Wood)
    d. Curation System Needs (from Terrence W Brady)
        i. PR 2181 implements per-run task parameters. Ready for review.
        ii. 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

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

Tim Donohue [3:02 PM]
@here: I'm running a bit late today...We do have a DevMtg now(ish)

Mark Wood [3:02 PM]
:slightly_smiling_face:

Tim Donohue [3:03 PM]
Ok, I'm fully here now (just got out of another mtg)
Our agenda for today is posted here: https://wiki.duraspace.org/display/DSPACE/DevMtg+2018-10-24
Let's do a quick roll call to see who @here is able to join us today

James Creel [3:04 PM]
I'm here.

Tim Donohue [3:04 PM]
Looking like we are a small three :wink: But, that's ok. We can see what progress we can make
The upcoming meetings are noted in the agenda...I won't repeat them, but will just let you look there for the
next DSpace 7, DSpace Entities & DSpace Dev Show & Tell mtgs
General updates on DSpace 7. I'll just note here that our Steering & Leadership groups are currently in
discussion around helping us promote/plan a "preview" (or alpha) release of DSpace 7 before the end of 2018
The discussion is really around what will be available/ready for the "preview" release. We'd like to get
something out there for folks to start playing with (and giving feedback on).
But, at this point, we can guarantee it won't have *every feature* you expect out of DSpace (and therefore more
features will continue to be added to DSpace 7 in early 2019)
I should have more to say on this next week (hopefully), or the following week. Obviously, with just Nov/Dec
to go, we're really just looking at where do we concentrate development in the next few months for the "biggest
benefit"
That's it on general DSpace 7 updates.
And, I don't have any updates on a 6.4 release. I imagine it'll happen at some point, but no immediate plans
(as everyone is concentrating more on 7 right now)
So, that brings us down to a list of PRs that I wanted to promote in this meeting -- simply because they are
PRs that are more around *updating our dependencies*....not the most exciting topic, but something that anyone
with knowledge of DSpace 6 should be able to help us approve & get merged.
First up, I have a PR to update all our webapps to Servlet Spec 3.1: https://github.com/DSpace/DSpace/pull/2235
It's a small PR overall, should be easy to review & test. I also realized today that (as a side-effect) likely
solves some older OAI-PMH related oddities that we were seeing on `master`, like https://jira.duraspace.org
/browse/DS-3806
Basically, as part of this dependency update I was forced to cleanup the OAI POM a bit more than others.

Mark Wood [3:14 PM]
It will be nice to align all of our webapp.s on one version, and a non-ancient one too!

Tim Donohue [3:14 PM]
Agreed. I would love it if I could get one more +1 on PR#2235, so that we can get this merged. (and thanks
@mwood for already giving your approval)
(that said, I know we are a small group today...so, I'm not sure if I'll find another +1 here or not, but I'll
keep trying to get this merged ASAP)

James Creel [3:15 PM]
I can take a look at 2235 tomorrow.

Tim Donohue [3:16 PM]
Thanks @jcreel256. I'd appreciate any feedback you can give!
Moving along, another dependency update from me: https://github.com/DSpace/DSpace/pull/2244   This updates our
Apache Commons Configuration to v2 (from v1), and Commons Lang to v3.
It "looks big", but 95% or more of the changes are changing package `import` statements (mostly for Commons
Lang actually). And, the good news here is this area of the code already had extensive Unit Tests (and I added
more in the PR). All those Unit Tests pass

Mark Wood [3:18 PM]
Indeed, reviewing goes quickly since almost all the files changed have only a single import statement altered.

Tim Donohue [3:19 PM]
While this doesn't seem "high priority" to upgrade, it'll let me fix another Spring Boot configuration issue in
the REST API v7: https://jira.duraspace.org/browse/DS-3492   (As this issue is dependent on upgrading Commons

```
Config to v2, which has better Spring integration)
So, again, I could use some code reviewers here.

Mark Wood [3:19 PM]
Snyk gets noisy when we let our dependencies go backlevel....

Tim Donohue [3:21 PM]
Snyk is very complainy about the version of Spring we are using across the board on `master`.  That's a known
issue, but it's blocked by some changes that are coming in future PRs.  It's mostly described in the latest
comment on this PR: https://jira.duraspace.org/browse/DS-3802   (Essentially, we are blocked from upgrading
Spring Boot / Spring until we fix an issue in our REST Code which will be fixed soon)
So, I'd request please ignore Snyk noise for now that relates to Spring or Spring Boot.  We will get these all
fixed prior to DSpace 7, but don't hold it against any current PRs

Mark Wood [3:22 PM]
OK
My point was that updating dependencies may be more interesting that it seems.

Tim Donohue [3:23 PM]
In any case, I could really use eyes on https://github.com/DSpace/DSpace/pull/2244 in the coming days.
@mwood: oh, yes, I agree...some of these dependencies need it badly, and it also helps with Java dependency
warnings we're now seeing from GitHub (which also are mostly about Spring/Spring Boot on `master`)
@mwood: would you have the opportunity to give PR#2244 an initial review in the next few days (or week)?  I get
the sense you looked at it, but maybe didn't finish up the review?

Mark Wood [3:25 PM]
That's about where I stand.  I want to do a proper review but haven't finished yet.

Tim Donohue [3:26 PM]
Ok, sounds good. I'd appreciate it whenever you could find some time.  I feel confident the unit tests are
ensuring the upgrade is a smooth one, but more eyes are welcome

Mark Wood [3:26 PM]
I feel that we want to get that one in soon and start breaking whatever is breakable, rather than let it
languish until the last  minute.

Tim Donohue [3:27 PM]
Next on this list of dependency updates.. the Log4j upgrade: https://github.com/DSpace/DSpace/pull/2241
@mwood: did you have any updates on log4j?  Any help you need on this new PR?

Mark Wood [3:27 PM]
Ah, I was hoping that 2235 would merge quickly so I could rebase on it, because that will help.  I can cherry-
pick it instead, and the duplicated commits will vanish when 2235 merges.

James Creel [3:28 PM]
Ok, so as not to inconvenience you, I will try to do a test of 2235 before I leave today
better start compiling

Tim Donohue [3:29 PM]
That sounds great, @jcreel256.  If you don't get to it, I can also push folks to review it in tomorrow's DSpace
7 Mtg (in the morning)
Either way, I think we can try and get 2235 merged quickly, @mwood, if that's of help here

Mark Wood [3:29 PM]
Right now the CLI is logging but I've not seen any log files created by the webapp.s.  Log4j2 should auto-start
in a Servlet 3+ webapp.
OK, thanks, that will help.

Tim Donohue [3:30 PM]
Oh, nice!  I didn't know that.  Could you add a comment about that to your PR, so that we know the two are
connected?

Mark Wood [3:30 PM]
Yes.

James Creel [3:30 PM]
It's been a little while since I built from the upstream master - just to speed me along, what would I need to
configure quickly?
dspace.cfg would suffice?

Tim Donohue [3:31 PM]
```

Yes, you basically just need database configs at the basic level.
upstream master is actually not much different from configuring DSpace 6. The main difference is you no longer
have the XMLUI or JSPUI... and have the new REST API instead. Beyond that, the configuration & build is the
same

James Creel [3:32 PM]
got it (edited)

Tim Donohue [3:33 PM]
One other dependency upgrade to mention -- which I left off this list accidentally: Upgrading Solr:
https://github.com/DSpace/DSpace/pull/2058

Mark Wood [3:33 PM]
If you want any user interface, you'll also need to build and install dspace-angular.

Tim Donohue [3:34 PM]
@mwood: I'll note here, I've pulled down your PR#2058 this morning, and I've made some headway on it. I can
possibly "steal away" this Solr Upgrade from you, if needed.
(I was initially just looking at where this PR#2058 stood, and then I realized your "blocker" was something I
knew how to fix...and now it's mostly working for me on a local branch)
"mostly working" means it compiles, passes *most* unit tests...with a few small stragglers

Mark Wood [3:35 PM]
Ah, I've let that one fall by the wayside, I see.
I need to make a list of all my TODO lists....

Tim Donohue [3:36 PM]
So, it's up to you here... if you have plenty of other tasks, I might be able to take this on (little by
little). Otherwise, if you want to continue this PR, I can push up a few commits to "fix" the main blocker you
were hitting (which was a Spring Boot + Solr conflict of sorts)
In the end, I'd just REALLY like to get Solr upgraded in DSpace 7. This is something I can put time into
(before Dec) if no one else can

Mark Wood [3:37 PM]
Let's start by pushing up your fix, and see how I go from there.

Tim Donohue [3:38 PM]
Ok, I can push up some fixes tomorrow to your PR (I've gotta logoff right after the meeting today)

Mark Wood [3:38 PM]
I should mention: I have it in mind to get Solr into a state that we can change the Solr server from a
dependency to a prerequisite.

James Creel [3:39 PM]
Like a separate installation?

Mark Wood [3:39 PM]
Yes.

Tim Donohue [3:39 PM]
@mwood: right, I have the sense that we'll *have* to do that. As of Solr 5, Solr server isn't even really
"embeddable"...so, it has to be a prerequisite as soon as we upgrade.

James Creel [3:40 PM]
Makes DSpace a little less "turnkey" perhaps, but I think it is a positive development. There could be some
scripts or something to ease deployment for some institutions.

Tim Donohue [3:40 PM]
We've been holding back on Solr 4 cause we have it embedded....as of Solr 5 it's a standalone server
application (not a webapp to be deployed). See https://jira.duraspace.org/browse/DS-2795

Mark Wood [3:41 PM]
Solr is dead simple to install. It's writing the core configurations that can become complicated. But we
would still supply that.

Tim Donohue [3:42 PM]
While we are on the topic of Solr, I'll note that upgrading Solr to version 7 (latest version) is also somewhat
dependent on this Solr Schema PR: https://github.com/DSpace/DSpace/pull/1810

In testing @mwood's PR this morning, I noticed some Unit Test failures immediately cause it seems Solr 7 is
more strict about checking against the Schema.

I'll add a note about that to PR#2058 once I push up my changes

Mark Wood [3:43 PM]
We need some schema updates anyway. We are still calling fields "integer" when they now contain UUIDs.

Tim Donohue [3:44 PM]
yep, exactly. And that's what our unit tests started saying when I ran it with Solr Client v7...it was
complaining about "Invalid integer: [some UUID]" all over the place
So, we'll need to streamline getting 1810 merged soon too (I know that's high on @terrywbrady's list, but he
couldn't make it today)
And, in fact, that was the next item on this agenda. Reviews/testing of PR #1810: https://github.com/DSpace
/DSpace/pull/1810   And, I know @terrywbrady would like to see this PR also backported to 6.x branch.
I'll admit, with 1810, I haven't had a chance to fully test this myself...but I *did* test the Schema changes
(in playing with the Solr upgrade in PR#2058) and they work

Mark Wood [3:47 PM]
That plays nicely with some local development which tripped over UUIDs.
I mean: I haven't tried 1810, but I made similar schema changes and it would be nice if they became stock code.

Tim Donohue [3:48 PM]
@mwood: any chance you have local indexes that you could test this upgrade script with?

Mark Wood [3:48 PM]
I reckon I could just copy the cores and try it.

Tim Donohue [3:49 PM]
Yes, if you have some older statistics (DSpace 5 or less), you should be able to copy them over to a dev server
and run this script to see if they get updated for DSpace 6+ (integer -> UUID)
I think that's all this 1810 is waiting on...a sanity test. I can give it a quick code review in the meantime

Mark Wood [3:50 PM]
OK, I will try it.

Tim Donohue [3:50 PM]
Thanks, @mwood
So, my plan for this meeting is to keep this sort of "running list of Java dependency update PRs" in our
Agenda. I think the group in these meetings could really help out DSpace 7 even by just reviewing/approving
these dependency updates.
(We've had too many dependencies which were locked down for so long cause of our old UIs, especially Cocoon.
We're now able to move a lot of these forward much faster
So, I'm thinking I might stop the meeting right about here. I have to leave promptly at the top of the hour.
But, before I do, is there any other topic anyone wants to raise to attention (for a brief 5 mins)?

Mark Wood [3:54 PM]
I'd like to reorganize the Curation doco.
It all sits in "Developing DSpace" which is surprising if you just want to know how to use existing tasks. I
think it should be split into "developing tasks" (which can stay there) and "using the curation system".
Probably the usage part should be under "Using DSpace", but would "System Administration" be better?

Tim Donohue [3:56 PM]
Makes sense to me. https://wiki.duraspace.org/display/DSDOC6x/Curation+System
Yes, I like "Using DSpace" and "System Administration" better
"Developing DSpace" should likely be a rather small section overall
And I've gotta run, unfortunately. Sorry. But, this direction sounds good to me

Mark Wood [3:57 PM]
I'm leaning toward Using DSpace. It really is aimed at content admin.s rather than sysadmin.s
Once I get the sections disentangled, the usage section can be moved easily if need be.
OK, I'll work on factoring the page.
I have to go, also. Thanks, all!

James Creel [3:59 PM]
Sure. The testing of 2235 looks like I won't finish today, but I'll be on it tomorrow as well.

Mark Wood [3:59 PM]
Thank you!