DevMtg 2019-07-17

Developers Meeting on Weds, July 17, 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, July 18 at 14:00 UTC.
- DSpace 7 Entities Working Group (2018-19): Next meeting is Tues, July 23 at 15:00 UTC.
- 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. Quick Updates from other meetings
   a. DSpace 7 Status Updates for this week (from DSpace 7 Working Group (2016-2023) or DSpace 7 Entities Working Group (2018-19))
   b. DSpace 6.x Status Updates for this week

      i. 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.
2. Report from LYRASIS Merger
   a. Merger completed as of July 1: https://duraspace.org/lyrasis-and-duraspace-complete-merger-members-and-community-benefit/
   b. **DSpace Public Events Calendar is being moved.** New iCal & HTML (New meetings/events can be added by Tim or Heather Greer Klein)
      i. The Old "DuraSpace Public Events" calendar will be deleted in the near future. If you use it, please switch to the new calendar above.
3. Ongoing Work
   a. Upgrading Solr Server for DSpace
      i. **n Solr**:   Unable to locate Jira server for this macro. It may be due to Application Link configuration.

         s only happen for **major** releases? Should it be configurable? Can we find a more precise trigger? When **do** we need to reindex?
      ii. ty core.   Unable to locate Jira server for this macro. It may be due to Application Link configuration.   Or shou
   b. DSpace l s (old) (Terrence W Brady)
      i. Update sequences on initialization

         1. https://github.com/DSpace/DSpace/pull/2362 - update sequences port
         2. https://github.com/DSpace/DSpace/pull/2361 - update sequences port
      ii. DSpace Launcher Dashboard - Deploy a PR on AWS for Testing
         1. There is a 2 minute video that illustrates this proposal.
4. For Discu
   a.   Unable to locate Jira server for this macro. It may be due to Application Link configuration.
   b.   ext" object had a separate DB connection. In DSpace 6+, each **thread** has a separate DB Connection (but Context objects may share a DB connection if they share the same thread).
5. 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 3:00 PM
@here: it's DevMtg time.  Here's our (late) agenda for today: https://wiki.duraspace.org/display/DSPACE
/DevMtg+2019-07-17

Let's do a quick roll call to see who is able to join
Terry Brady 3:01 PM
hello
Mark Wood 3:01 PM
Hi
James Creel 3:01 PM
Scouring a database for language qualifiers, but I'll be listening in.
Tim Donohue 3:02 PM
Ok, we'll go ahead and get started with quick updates from other meetings.
On the DSpace 7 side, we have two simultaneous working groups (DSpace 7 Entities & DSpace 7).  Both are very
active and meeting weekly.  I don't have any major news to report from either, but others are welcome to follow
along from meeting notes, jump in a meeting, or follow on Slack
The general news though is both groups are active, lots of work going on. Nothing major to report at this time,
just a lot of day by day / week by week feature development, etc
But, I'll pause here if anyone has any questions, or anything else they want to mention regarding these groups?
Not hearing anything. If questions come up, feel free to ask on #dev  (or in a more specific slack channel)
On the DSpace 6 side of things, no progress to note.  While PRs are still occasionally coming in, 6.4 isn't in
progress yet (still waiting on someone to lead that release)
So, that's it for the quick updates on v6 and v7.  Obviously, as always, if you'd like to be more involved,
please get in touch or just join a meeting. There's plenty of space/opportunities for involvement
As topic #2 today, I adding in a brief note on the LYRASIS merger (as I realized I hadn't mentioned anything in
this meeting yet!)
If you haven't seen the news, DuraSpace and LYRASIS are fully merged as of July 1: https://duraspace.org
/lyrasis-and-duraspace-complete-merger-members-and-community-benefit/
Duraspace.orgDuraspace.org
LYRASIS and DuraSpace Complete Merger—Members and Community Benefit - Duraspace.org
CONTACT Meg McCroskey Blum, Director of Marketing & Communications, meg.blum@lyrasis.org, 800.999.8558 x2951,
Skype: meg.lyrasis LYRASIS and DuraSpace are pleased to announce the merger of their two leading 501 C3 not-for-
profits was completed on July 1, 2019, Members will participate in developing new scalable technologies, shared
innovation opportunities, and high value – fairly priced services across the global landscape of... Read more »
Jul 9th
This means that I now work for LYRASIS (as do all former DuraSpace staff). My job/role/responsibilities have
not changed though, so you won't really see any difference on the DSpace front.
The only recent change (more for me than anyone else) is I've had a lot more meetings as of late (a lot of
getting to know each other, and all the open source projects under the LYRASIS umbrella).
But, I expect that to calm down in the coming weeks :wink:
```

All that said, as noted in the agenda, I have been working on moving the old "DuraSpace Public Events Calendar" over from Google Calendar to MS Outlook (LYRASIS uses MS tools).
So, I've got a brand new "DSpace Public Events Calendar" here: https://outlook.office365.com/owa/calendar /88006a3cc8c64f30a44b78f8dceb7156@lyrasis.org/972e73d7e8cc4d5096e913f1df94ab3116487484544154616150/calendar.html
And the iCal version is at: https://outlook.office365.com/owa/calendar/88006a3cc8c64f30a44b78f8dceb7156@lyrasis. org/972e73d7e8cc4d5096e913f1df94ab3116487484544154616150/calendar.ics
I'd like to find a volunteer (or two) to try out this new calendar, as we are looking to turn off the old DuraSpace Public Events calendar in the near future.
(So, if you used to use the old DuraSpace one for meeting reminders, you'll want to use this new one instead)
Mark Wood 3:12 PM
Fortunately we also are afflicted with Exchange, so I already have that plugin for Thunderbird/Lightning working as well as it ever does.
Tim Donohue 3:13 PM
@mwood: I'd definitely appreciate it if you could try this out then. This is my first time creating a Public Outlook Calendar...I think it's working (from my own tests), but would love someone to verify before I post this more publicly (on Wiki, etc)
Mark Wood 3:13 PM
OK, I'll try it.
Tim Donohue 3:13 PM
Thanks!
Terry Brady 3:13 PM
I am able to subscribe to it in Google Calendar
:+1:
1

Tim Donohue 3:14 PM
Good :slightly_smiling_face: In that case, it's likely working. But, let me know what you all find out as you use it.
As a sidenote, as many of you may have seen...my email address similarly has changed. My old @duraspace.org email will likely work for a good long while (and it redirects all mail), but expect to see emails from me coming from a new email address. I'd also recommend updating your address books
That's all the updates I have on the merger front. Are there any questions about the merger?
(Oh, and I should mention the plans are the Wiki & JIRA & Slack & all the other tools we are using should not be changing. The URLs might eventually change from duraspace.org to lyrasis.org, but no timelines on when that would happen)
:clap:
1

Mark Wood 3:17 PM
Thunderbird Exchange plugin worked just fine (so far) using the .ics link.
:+1:
1

Tim Donohue 3:19 PM
Ok, sounds like no questions. If anything comes up, let me know. And, I'll be sure to let you know if anything else comes up to be aware of. As noted though, I don't think you'll see any major changes...just eventually some minor ones (like URL changes)
Moving along to other topics now. We are into the "Ongoing Work" updates
@mwood: any updates to note on the Solr upgrade process this week? https://wiki.duraspace.org/display/DSPACE /Upgrading+Solr+Server+for+DSpace
Mark Wood 3:21 PM
Not much to say. Still putting together a testbed for upgrading from pre-7. You may have noticed a drizzle of patches to the command-line tools to support export/import from one of our live instances. This afternoon I made a gadget to remove the entire content of my test instance since I imported the structure without preserving the Handles, oops.
I should be fairly close now to actually testing the upgrade. Sorry it's taking so long.
Tim Donohue 3:23 PM
no worries, I know other things come up. Luckily, this hasn't been a "high priority" item (in that we need to rush it). Obviously, we want it as soon as we can get it, but DSpace 7 beta is still a ways off
So, as always, just let us know when you need help from others (testers, reviewers, etc)
Mark Wood 3:24 PM
Will do.
Tim Donohue 3:25 PM
Moving along, @terrywbrady: is there any updates you want to add about DSpace + Docker this week? https://wiki.duraspace.org/display/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals
Terry Brady 3:25 PM
Nothing new. The 2 PR's are still open.
Tim and I will meet with some Lyrasis folks to talk about the DSpace Launcher Dashboard to see if the concept could be supported.
Mark Wood 3:26 PM
They are still in my to-test list.

```
:+1:
2
```

Tim Donohue 3:27 PM
Sounds good, and yes, Terry will be showing off the DSpace Launcher Dashboard to a few folks in LYRASIS (basically ex-DuraSpace tech folks) to get early thoughts. If there's interest, we can bubble it up to other projects in LYRASIS too

Terry Brady 3:29 PM
I hope to do some kubernetes exploration in the next couple weeks so that I can offer some thoughts on how we might be able to use it for DSpace deployments in the future. I successfully deployed DSpace to Google Cloud (in Docker) yesterday.

```
:+1:
1
```

Tim Donohue 3:29 PM
Very cool. Glad you still are finding time to dig around in this area. It's exciting to hear about
Ok, so, moving along now...
I kept this next topic on the agenda from last week "Brainstorming how to improve DSpace database usage (now that we use Hibernate)"   Is there more to say on this topic?
I read the notes from last week. It seemed like there was general consensus it'd be good to dig more here... but no real next steps that I heard. Is the next step "wait until someone has time"?

Mark Wood 3:32 PM
Nothing new there either. I've begun looking at pulling the Session out of Context, but haven't gone far.

James Creel 3:32 PM
Theres a card on our local backlog to try to duplicate performance issues we think we have
but no sprint scheduled presently

Tim Donohue 3:32 PM
Thanks @jcreel256 for that note. Definitely interested in hearing about performance issues that we can duplicate.

```
:+1:
1
```

On that note, I'll mention this same topic (performance testing) is coming up in tomorrow's DSpace 7 meeting. @cwilper has run some early performance tests against DSpace 7 specifically an will be giving an overview of the results he found: https://cwilper.github.io/dspace-perftest/
dspace-perftest
DSpace Performance Testing
Performance Testing for DSpace

Mark Wood 3:34 PM
I will be all ears.

Tim Donohue 3:35 PM
The results are generally that we do have some performance issues we've found...but need to dig more into the exact cause(s). It's also possible some of these performance issues are in both DSpace 6 and 7 (I honestly don't know for certain, but some of the results imply certain backend activities are "slower" than expected)

Terry Brady 3:35 PM
If we are able to assemble reusable performance testing datasets, I would be interested in documenting those datasets for re-use.
https://github.com/DSpace-Labs/AIP-Files

Tim Donohue 3:36 PM
In any case, the reason I mention this in this meeting as well is that there may be a good overlapping discussion forming... and if we find folks interested in helping dig more (collaboratively) on performance issues (that can be replicated), that'd be wonderful. I think we're more likely to squash these quickly if we get a few folks working together
(So if you are interested, join the discussion tomorrow...or if you cannot, get in touch. I think this will be an ongoing discussion for some time as we dig into the results, etc)

Terry Brady 3:38 PM
I will miss the meeting tomorrow, but I will be interested.

Tim Donohue 3:38 PM
@terrywbrady: Per your point on reusable data...my understanding is Chris used some "dummy data", but he posted his process here: https://wiki.duraspace.org/display/DSPACE/DSpace+7+Performance+Testing
I think Chris has a script to create larger amounts of "fake" data in DSpace... I need to remember to ask him where it is though, so others can use it

Mark Wood 3:39 PM
He does, and so do I.

Terry Brady 3:39 PM
That is awesome. I would like to link that to the repo I listed above.

Tim Donohue 3:40 PM
We should share those scripts somewhere :slightly_smiling_face:

Mark Wood 3:40 PM
I think both Github repo.s were linked somewhere in Slack just today, or maybe yesterday.

Tim Donohue 3:41 PM

@terrywbrady: To clarify my point...I'm not sure if these two "data sets" are serving the same purpose. The AIP-Files are "real looking" data sets for user testing, while these bulk creation scripts just create a lot of dummy content (that is not pretty for a human) that works well for performance testing.
While both are needed, we might not be able to merge the two ideas into one
Mark Wood 3:42 PM
Yup, yesterday and today in #angular-ui
Tim Donohue 3:42 PM
aha, thanks @mwood. I'm going to link those scripts into the DSpace 7 Performance Testing wiki page
Here's Chris's scripts: https://github.com/cwilper/dsogen
And I just linked them to that DSpace 7 Performance Testing page
So, we are at the end of the agenda. Are there other topics from anyone?
Mark Wood 3:46 PM
I'll just mention that I have a little project for those weekends when the weather is too awful to go out: refactoring the password authentication to use the PBKDF2 hash support in the JDK .
Tim Donohue 3:47 PM
Sounds interesting. I admit, I don't know much about PBKDF2, other than it's "more secure" for passwords, etc
But, cause of that, it sounds like a good idea :slightly_smiling_face:
Mark Wood 3:48 PM
The formal Key Derivation Functions are supposed to be deliberately slow memory hogs, to foil brute-force attacks.
It's probably not a burning need, but we ought to keep up with recent developments in security and it sounded like fun.
Tim Donohue 3:49 PM
yep, agreed. Not super high priority, but sounds like a nice to have eventually
Any other updates / topics that anyone would like to share?
Ok...doesn't seem like anyone has anything else. So, let's close up today's meeting 10 minutes early
Have a great rest of the week, and again feel free to join the DSpace 7 meeting tomorrow if you want to hear about / discussion performance testing of DSpace 7!
Thanks all!
Mark Wood 3:51 PM
'bye, all.
Terry Brady 3:51 PM
bye