

DevMtg 2018-09-19

Developers Meeting on Weds, September 19, 2018

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)
 - 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 is tomorrow, Thurs, September 20 at 14:00 UTC
 - **Signup for our next Community Sprint (Oct 1-12): [DSpace 7 Community Sprints](#)**
- [DSpace 7 Entities Working Group \(2018-19\)](#): Next meeting is Tues, Sept 25 at 14:00 UTC
 - Latest meeting (including video recording): [2018-09-11 DSpace 7 Entities WG Meeting](#)
 - Early prototype writeup / docs: <https://docs.google.com/document/d/1wEmHirFzrY3qgGtRr2YBQwGOvH1luTVGmxDIdnqvwxM/edit>
- [DSpace Developer Show and Tell Meetings](#): Looking for more meeting ideas!

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 Development Status spreadsheet](#)
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 Release 5.10 Status](#)
4. Encouraging / promoting code contributions (PR creation) and code reviews ([Tim Donohue](#))
 - a. "DSpace's Top GitHub Contributors" site: <https://tdonohue.github.io/top-contributors/>
 - b. Code at: <https://github.com/tdonohue/top-contributors>
 - c. Tim is looking for feedback on these statistics, and whether there are other useful ways to credit efforts of individual developers
5. Brainstorms / ideas (*Any quick updates to report?*)
 - a. [Bulk Operations Support Enhancements](#) (from [Mark H. Wood](#))
 - b. [Curation System Needs](#) (from [Terrence W Brady](#))
 - i. [PR 2181](#) implements per-run task parameters. Ready for review.
 - ii. [PR 2180](#) improves reporting. Needs a little more testing
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/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

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

Tim Donohue [9:52 AM]

@here: It's almost DevMtg time. Agenda is at: <https://wiki.duraspace.org/display/DSPACE/DevMtg+2018-09-19>

Tim Donohue [10:00 AM]

@here: It's now DevMtg time! The agenda is posted above. Let's do a brief roll call to see who is able to join us today.

Mark Wood [10:01 AM]

Hi

Pablo Prieto [10:01 AM]

Hi

Terry Brady [10:02 AM]

hello

Tim Donohue [10:02 AM]

Hi all, looks like we have (at least) a small quorum.

So, let's get started. First a few quick reminders...

We have our next DSpace 7 Community Sprint coming up in a few short weeks (Oct 1-12). Signups are still open at: <https://wiki.duraspace.org/display/DSPACE/DSpace+7+Community+Sprints#DSpace7CommunitySprints-ParticipantSignups>

Currently 3 developers signed up, but we'd love to have more!

The next DSpace 7 Entities WG meeting is next Tues (Sept 25 at 14UTC): <https://wiki.duraspace.org/display/DSPACE/DSpace+7+Entities+Working+Group#DSpace7EntitiesWorkingGroup-NextMeeting>

Dev Show & Tell is still looking for future topics. So, there's always the opportunity to bring ideas for a Dev Show & Tell to one of these meetings

Terry Brady [10:05 AM]

I am still figuring out my availability. I hope to participate in the Sprint in some form.

Tim Donohue [10:05 AM]

I think that's it for quick reminders.

@terrywbrady: that'd be great. If you even have very rough estimate, feel free to signup with a range (e.g. 10-20% or similar). It'd be good to add your name to the list if you know you'll be contributing

Generally, I find the more names on the list, the more likely others also "jump in" and join up. :wink:

Pablo Prieto [10:07 AM]

I'm in the same status as @terrywbrady

Tim Donohue [10:08 AM]

Moving on into agenda topics. Other than the reminders above, I don't have any specific DSpace 7 updates to share today. Same goes for DSpace 6.x... so, I think we can skip over #1 and #2 on the agenda, unless anyone has comments/questions to add on either.

@Pablo Prieto: Sounds good, I'd also recommend adding your name then with a rough estimate. It helps the sprint coaches to better understand how much time we need to "block off" if we know there are going to be 5-6 participants, instead of just 3. :slightly_smiling_face:

Terry Brady [10:10 AM]

@tdonohue, have you had a chance to reach out to the new folks who have joined Slack and who attended the last Show and Tell meeting?

Tim Donohue [10:10 AM]

So, if you plan to participate in the sprint (in any form), it'd be helpful to me if you can add your name (even if you haven't figured out the exact percentage of participation yet)

Terry Brady [10:10 AM]

I will add my name with a TBD on availability.

Tim Donohue [10:11 AM]

@terrywbrady: No, I didn't have their contact info. I dug around briefly, but at the time, I don't think everyone had even joined Slack.

I will be sending general reminders on Slack though this week (likely today or tomorrow) as well as via email. So, those general reminders will go out again soon

Moving along here for now.... on to topic #3, DSpace 5.10 Release updates: <https://wiki.duraspace.org/display/DSPACE/DSpace+Release+5.10+Status>

@terrywbrady: is there any updates to share on 5.10? Should we be penciling in a possible release date?

Terry Brady [10:14 AM]

I am waiting to get some time with @pbecker to test the RDF service (and then to document how to test that service going forward). I believe it also had a reference to one of the jar files that was triggering issues. If we are unable to set up a time to meet, perhaps I will just move forward with the release. How useful would it be to you all to have Dockerfiles to support various versions of Java/Tomcat? Do we have any intention to go to Java9/Tomcat9 soon?

Tim Donohue [10:16 AM]

I know @pbecker was very busy last week (German User Group meeting). He may still be catching up. Hopefully he will pop online today/tomorrow and give us an update.

I agree with you though that I think we should move forward soon. My suspicions are that the primary issues have been solved.

Terry Brady [10:16 AM]

I ask because I needed to create Java7 Dockerfiles to verify this release. I read the DuraSpace newsletter. It sounds like his user group meeting was a great success!

Tim Donohue [10:18 AM]

I'm not sure whether the effort involved in maintaining separate Dockerfiles per Java release is worth it. But, I'm fine either way.

As for the Java 9 question, that's more a question for DSpace 7 team. We never update the Java versions of *old* releases... so DSpace 6 and below will never be released with Java 9.

Terry Brady [10:19 AM]

If you hear rumors of an update, let's chat about the Docker implications. Until the recent bugs came up, I did not realize that DSpace 5 was supported on Java 8.

Tim Donohue [10:20 AM]

That doesn't mean that DSpace 6 and below may not work with Java 9...but, DSpace 6 installation docs specifically state Java 7 or 8. We never add "Java x or above" DSpace 5 is actually not specifically supported on Java 8. But it works on Java 8. We made a mistake of accidentally adding in a PostgreSQL driver for Java 8 in 5.9. Java dependencies/support is frankly all a bit difficult. It's hard when we support 3 releases at once...and some of those releases now use EOL Java versions.

Terry Brady [10:22 AM]

And the Java releases are coming out faster than the DSpace releases!

Tim Donohue [10:22 AM]

yes, exactly. And that seems like it may only ramp up further. Java has been talking about new releases every *6 months*

Mark Wood [10:23 AM]

Indeed. Java 9 is already dead. Oracle is shipping Java SE 8 and 10.

Tim Donohue [10:24 AM]

So, this will be an ongoing topic. Maybe it is worth looking at minimally allowing our Dockerfiles to be *configurable* for the Java version? (if that's possible)

Terry Brady [10:25 AM]

I did some work on this for Java 7, so we can use that as an example. I think we need to either (1)recommend 1 version of Java for a DSpace release and code that into the Dockerfile or (2)support N versions of Java for a DSpace release and then publish an image that contains the JDK version and the DSpace version.

Mark Wood [10:26 AM]

My guess is that other stuff was held up so long by Jigsaw that they need to catch up, and that the schedule may be lengthened after a few major releases.

Tim Donohue [10:27 AM]

@terrywbrady: well, technically, we could just follow the Installation instructions per release. 4.x says Java 7 (only). 5.x says Java 7 (only), 6.x says Java 7 or Java 8. So, that really limits the number of Dockerfiles we'd need. We do *know* that 5.x works on Java 8 (so that's optional), but it's not "officially supported" as such

Terry Brady [10:29 AM]

That probably makes sense. The 6x is interesting since we do mention 2 options. I question if 6x really can run on Java 7. I suspect that some 8x JAR files have already bled into the distribution.

Tim Donohue [10:29 AM]

@terrywbrady: could be. I don't know for certain either.

Terry Brady [10:29 AM]

I will add a ticket and reference our Docker repo to capture this conversation. This will require some more thought.

Tim Donohue [10:29 AM]

In any case, it sounds like we have some ideas on moving forward here.

On the 5.10 timeline side of things, I'd recommend starting to look at potential dates -- so we can finalize this release *next week*, if not sooner. Hopefully that gives @pbecker enough time to weigh in.

Any last notes/comments on this topic (of 5.10 / Dockerfile Java versions)?

Not hearing any (or seeing anyone typing)...moving along

As announced on #dev yesterday, I've made a bit of a recent "breakthrough" on promoting developer contributions in GitHub (PR creation / reviews): <https://tdonohue.github.io/top-contributors/>

tdonohue.github.io

DSpace's Top GitHub Contributors

Ranking DSpace GitHub contributors, since 2018

Codebase at: <https://github.com/tdonohue/top-contributors>

Mark Wood [10:32 AM]

It looks sharp.

Tim Donohue [10:32 AM]

I'd really love feedback on this work

Ideally, if others like this approach, I'll sign up to "maintain it" (i.e. update at the end of each month).

I'd also recommend we move it over into the main DSpace or DSpace-Labs area

Terry Brady [10:33 AM]

It looks awesome! It adds some of the gamification from Stack Overflow.

Tim Donohue [10:35 AM]

This morning, I'll note that I also realized that I likely can pull out *Company* information from user accounts (to credit your organization, if it's listed in your GitHub profile). So, that might be a minor update to the awards tables: <https://tdonohue.github.io/top-contributors/2018/09/01/Top-August-Contributors.html>

DSpace's Top GitHub Contributors

Top August Contributors

The Awards for August 2018 go to...

Aug 31st

I suspect that update would only take me about 30-60mins to push out. So, it might be there later today even.

Do you all have any thoughts on where this thing should "live"? Should I move it to DSpace-Labs? Or even the main DSpace repo (so that the URL would be dspace.github.io/top-contributors)?

Terry Brady [10:37 AM]

This seems like a DSpace-Labs thing. You might be able to make it a "pinned" repo on the DSpace org page.

Mark Wood [10:37 AM]

If it's working, stable, and "official" then it should go to DSpace.

Otherwise DSpace-Labs would be good.

Terry Brady [10:39 AM]

Pasted image at 2018-09-19, 8:39 AM

Tim Donohue [10:39 AM]

It is working, stable and as "official" as it can be. Though, to make it fully "official", I can run this by Steering for final approval.

I'm not sure that I can "pin" a repository in another org (i.e. "DSpace-Labs"). It looks like I can only "pin" a repository from "DSpace" org

Terry Brady [10:41 AM]

As a user, I can pin from my orgs. I suppose that it makes sense that an org is restricted.

I am fine if either org is chosen.

Mark Wood [10:41 AM]

wonders if this thing can be generalized and made into one of those badges. Maybe every project on Github would want it.

Tim Donohue [10:42 AM]

How about I get approval here from Steering & then move this over to "DSpace" org in GitHub. I can *exclude* it from any of the actual statistics that it tracks (obviously, it'd be weird to see PRs to "top-contributors"

showing up in the stats) -- though I may not even manage it via PRs, we'll see.

@mwood: yes, I have already shared this with Fedora & VIVO staff as well. It likely could be generalizable. A certain portion of it is still a bit "manual" (in that you have to manually update/run a script to pull down /parse the data). but we might be able to find ways to automate further in future.

Mark Wood [10:44 AM]

Cool

Tim Donohue [10:45 AM]

In any case, thanks for all the feedback here! Glad to hear that you all see this as useful too! If anyone has ideas for additional "stats" to try to track, let me know.

I'll take this to our Steering Group and let you all know what they think (though I suspect they will also be in favor)

With that, I'll move along to other topics for today

Any updates on topic #5 brainstorm? Bulk Operations (from @mwood) or Curation System needs (from @terrywbrady)

Mark Wood [10:47 AM]

Nothing on bulk this week.

I was hoping to reach closure on just where workflow-triggered tasks ought to write their reports.

Tim Donohue [10:50 AM]

@mwood: understood. I'm still "ok" with workflow-triggered tasks using logs (just that if they do use logs, it should use log4j). But, I don't use these tasks as heavily.

I will also note that I wonder if we are trying to overthink this (without an exact use case to build to).

Should we just make a smaller step in the right direction here?

For context (for others), the PR I'm referencing here is this one: <https://github.com/DSpace/DSpace/pull/2180>

Mark Wood [10:51 AM]

I agree that no code should be independently writing to /logs

I think now that considering these reports to be log material was a mistake, but I've cooked up two pluggable report serializers (log and file) since it appears that opinions are diverse.

Terry Brady [10:52 AM]

I shared some details on my use case. Did my use case make sense to you all?

Mark Wood [10:53 AM]

Yes. I responded in the PR. <https://github.com/DSpace/DSpace/pull/2180>

Tim Donohue [10:54 AM]

@terrywbrady: Parts of it did, but I don't know that I understood the types of report that is most useful to this use case. Are you looking for a log file? A CSV report? An email notification? etc etc

Terry Brady [10:55 AM]

My gut says that an HTML doc or HTML fragment would be useful. If the file is saved to a retrievable location, the admin UI could open the doc in a new tab.

That doc could create headers and links appropriate to the task being performed. If no links are needed, the HTML could be simple text.

Mark Wood [10:56 AM]

So, is a file per run, perhaps in /reports, usable?

Terry Brady [10:56 AM]

Yes. I think that would work well.

Tim Donohue [10:57 AM]

I guess my general point here is I'd love more detailed "user stories" or "use cases" that can provide us hints on what type(s) of reporting is most useful. I'm not sure if it's always the same. E.g. while HTML is nice for Admin UI viewing in a browser, it's very "static" and hard to parse / do analysis on (whereas CSV is easier to parse / do analysis)

Mark Wood [10:58 AM]

Not to mention that the admin. running a task on the console might not appreciate all the markup.

A task could define a parameter to tell it what kind of output you want for this run.

Terry Brady [10:59 AM]

In my use case, I imagine that I would be generating a TODO list of cleanup tasks for a repository manager. But, there are other instance where I could imagine that we want to provide a CSV file for a bulk metadata update run.

Tim Donohue [10:59 AM]

@mwood: yes, maybe that's the best option, if it isn't too hard to make the export format configurable.

Terry Brady [11:00 AM]

Would it be useful to schedule a 30 minute conversation to dig deeper on these use cases?

Mark Wood [11:00 AM]

Perhaps.

Terry Brady [11:00 AM]

I am excited to see some traction on this, so thank you @mwood!

Mark Wood [11:00 AM]

You are welcome. This is an interesting problem.

Tim Donohue [11:01 AM]

Should this be a "Dev show & tell" of sorts?? It's less a "Show & Tell", but more a special "Developer Discussion" meeting.

realizes we are at the top of the hour here.

Mark Wood [11:02 AM]

Would it be useful to make the report saving pluggable, so you can write what you need? If nobody objects to the concept, I can push that up pretty quickly for inspection.

Tim Donohue [11:02 AM]

That concept seems reasonable to me... we could start with just a handful of report formats (HTML, CSV, maybe plaintext for CLI)

Terry Brady [11:03 AM]

That sounds good to me as well.

Mark Wood [11:03 AM]

I have untested "write to a file as-is" and "gather up lines and log each one" written up now.

Tim Donohue [11:03 AM]

Ok, so, I'm gonna have to head out here. It sounds like we have a general direction for next steps. Should we just make this a topic of discussion for next week's meeting?

Mark Wood [11:04 AM]

That sounds like a plan.

Tim Donohue [11:04 AM]

Ok, will do. I'll start up an agenda for next week right away then, and bump this up in the list.

Thanks for the discussion today, all! We'll close out today's meeting. Have a good rest of the week!

Mark Wood [11:05 AM]

I'll go ahead and push up the plugin work.

Terry Brady [11:05 AM]

Have a good week!

Mark Wood [11:05 AM]

Thanks, all.

Pablo Prieto [11:06 AM]

Thanks!