DevMtg 2019-03-20

Developers Meeting on Weds, March 20, 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, March 21 at 15:00 UTC. Agenda: 2019-03-21 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. (*Ongoing Topic*) DSpace 7 Status Updates for this week (from DSpace 7 Working Group (2016-2023))
    a. **Configurable Entities PRs** (Part 1 - Architecture, Configuration, Display, Search/Browse):
        i. REST Contract PR: https://github.com/DSpace/Rest7Contract/pull/57
        ii. REST API (v7) PR: https://github.com/DSpace/DSpace/pull/2376
        iii. Angular UI PR: https://github.com/DSpace/dspace-angular/pull/372
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 (Mark H. Wood)
    a. PR https://github.com/DSpace/DSpace/pull/2058
    b. Docker configuration for external Solr
        i. https://github.com/Georgetown-University-Libraries/DSpace/commit/7115173d61776dd2455690518f5c9809cd0f28d4
            1. The Dockerfile creates a new solr instance with 4 cores. It then overlays the schema and config changes in PR 2058.
            2. I attempted to create my branch so that I could create a PR back to Mark's branch, but some other changes from master seem to be showing up if I create a PR.
        ii. This will need a small change to our docker compose files to invoke the external solr service. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/79
4. DSpace Backend as One Webapp (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)
5. DSpace Docker and Cloud Deployment Goals (old) (Terrence W Brady)
    a. Passing environment variables to Docker
        i. https://github.com/DSpace/DSpace/pull/2373
        ii. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/97
        iii. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/98 (adds Sword overrides)
    b. Creating default AIP dataset for DSpace 7 docker load
        i. Tim shared a link to the entities WG dataset. This dataset contains no bitstreams. How should we handle this for the AIP's?
        ii. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/95
    c. 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
    d. 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
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. Bulk Operations Support Enhancements (from Mark H. Wood)
    c. Curation System Needs (from Terrence W Brady)
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
    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 [10:01 AM]
@here: It's DevMtg time.  Our agenda is at https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-03-20
Let's do a quick roll call to see who's able to join today

Mark Wood [10:01 AM]
Here.

James Creel [10:01 AM]
I'm prepping for a demo, but keeping this window open

Pascal Becker [10:02 AM]
Have to excuse myself today. Sorry

Tim Donohue [10:04 AM]
Ok, so it seems like a very small group today
I guess we'll get started with general updates though and see where it brings us
First up, on the DSpace 7 side, I think the major update is that we now have (rather large) PRs for
*Configurable Entities*. These are now open for general feedback, after having already gone through feedback
/approval in the DSpace 7 Entities Working Group
* REST Contract PR: https://github.com/DSpace/Rest7Contract/pull/57
* REST API (v7) PR: https://github.com/DSpace/DSpace/pull/2376
* Angular UI PR: https://github.com/DSpace/dspace-angular/pull/372
The REST Contract PR is the smallest (and a good place to start). The other two are very large, but have a lot
of background material to help you along, etc. As noted, all of this code has had a first stage of reviews,
If there are any questions or comments on this effort though, or how best to provide feedback, please do get in
touch (either directly, in a meeting like this, or in #dev channel)
On the DSpace 7 side, I'll also note that we have a list of PRs that are *required* for the Preview Release.
Most of those have now moved into the "Review stage" and the others should move there soon.
The list of those PRs are identified by blue stars in the "Current Work" section: https://wiki.duraspace.org
/display/DSPACE/2019-03-21+DSpace+7+Working+Group+Meeting#id-2019-03-21DSpace7WorkingGroupMeeting-CurrentWork
Overall, this means that we are running a bit behind on the Preview Release (hoped it'd be done by end of
March), but I feel like we seem "on track" for an early-to-mid April Preview Release.  This is all dependent
though on getting timely reviews for those "blue star" PRs though
That's it for DSpace 7 updates, unless there are any questions?
I don't have any updates on DSpace 6, unless anyone else does? (Will pause a moment)
Moving right along then..
@mwood: any updates / feedback you want to share on the ongoing Solr upgrade effort? https://github.com/DSpace
/DSpace/pull/2058

Mark Wood [10:15 AM]
```

2058 still stands at +2, with no comments in the last 9 days.
Work on the authority dump/restore is dependent on that PR so I've been working on other stuff.
(I did spend some time learning about the authority feature, so I can actually have something to test with.)

Tim Donohue [10:18 AM]
Did we close the loop on all the outstanding review tasks for PR#2058? Just wondering if we are ready to merge it, or if you are aware of anything outstanding? (I apologize, I should recall this, but it's been a busy week, and I was on holiday over the weekend)

Mark Wood [10:18 AM]
I should recall it too, but let me take a quick look.

Tim Donohue [10:19 AM]
If we need to, we could just tell the DSpace 7 team tomorrow: "Hey, we are merging PR#2058. Tell us now if you object".

Mark Wood [10:19 AM]
I think that the only issue was with upgrading old cores, which was not addressed in 2058 and looks like we can't do that anyway except by rebuilding them.
Yes, let's do that.

Tim Donohue [10:20 AM]
Ok, I'll write myself a note. But, please remind me in the "Standup" portion of tomorrow's meeting, as well

Mark Wood [10:20 AM]
OK.
Noted in passing: the documentation on how to use authority control is "TBD". There's a link to the original proposal, so I can probably figure it out anyway....

Tim Donohue [10:21 AM]
Thanks for the other updates. I'd recommend that we consider PR#2058 "finished", but we can verify 100% tomorrow. That should free you up to look more at the authority dump/restore more in the coming week

Mark Wood [10:22 AM]
That sounds well.

Tim Donohue [10:22 AM]
Yes, I've seen that unfortunate TBD note on authority control. If you are able to document anything you learn, that would be useful. It's an area of the code that I admit I don't even know very well yet

Mark Wood [10:22 AM]
I will try to capture what I work out.

Tim Donohue [10:22 AM]
thanks!

Mark Wood [10:23 AM]
That's all I have on Solr just now, unless there are questions.

Tim Donohue [10:23 AM]
Nope, none right now. It all sounds like it's moving in the right direction to me. We just need to get that first PR finally merged
Ok, moving right along now. I don't have any significant updates on the "DSpace one webapp" backend to share today.
I'll just note that the PR is still up for review. So, reviews are welcome! https://github.com/DSpace/DSpace/pull/2265
I'd really like to get this first PR merged in the coming few weeks (if possible), as it'll help me also move forward with renaming things, etc

Mark Wood [10:25 AM]
Yes, I promised to review, now I need to do that.

Tim Donohue [10:26 AM]
I see @terrywbrady added a review, which I can respond to. His only questions seem to relate to (known) configuration complications (which will be cleaned up in the follow-up "renaming" PR)... and with the generic Error Page that is already returned by `/spring-rest`
Thanks Terry for the review...I think some of that is a bit "out of scope" for this PR, but I'll add notes on how this will get cleaned up in the next PR.
In any case, I had no other updates on that PR to share. So, we can move right along
I'm not sure @terrywbrady is available today (I think he has a conflict)...so, I don't have anything to add about the DSpace + Docker updates, other than to note that Terry has added several PR links to the agenda

Mark Wood [10:30 AM]
I became sufficiently irritated with the environment problem that I'm trying to find my way into the Docker images to see where the shell is getting involved. That's the trouble spot. If we could just carry these properties straight from 'docker' to wherever they are consumed, we wouldn't need to translate them.
So far I'm just stumbling around, learning enough about the organization of Docker to find the right image to inspect.

Tim Donohue [10:31 AM]
Yes, the environment variable translation stuff is turning into a pain. But, honestly, I don't know if there's any easier way around it. I agree with Terry's latest comment on the ongoing PR though: https://github.com/DSpace/DSpace/pull/2373#issuecomment-474821975

Mark Wood [10:32 AM]
Yes, when I saw the proposal my first thought was "shorter!"

Tim Donohue [10:32 AM]
We need to keep these configurations named how they are (for backwards compatibility). If Docker is being a pain, then we need to just come up with a translation strategy we can "be ok with"

DSpaceSlackBot (IRC) APP [10:32 AM]
misilot has quit the IRC channel

Tim Donohue [10:33 AM]
Honestly, I don't think these environment variables being long is that big of a deal. They are just to get around Docker issues, and likely most people using Docker aren't even going to notice them (as they seem to be about initializing your Docker environment)

Mark Wood [10:34 AM]
I've looked over Pascal's PR and it looks right to me, but I am holding off reviewing until the environment thing settles down.
True, we should be able to address the exact translations and move on quickly.

Tim Donohue [10:36 AM]
Sounds like we agree in general here. It's annoying, but we can live with it for now. We just need to make a final decision and go with it. (We can always change it in the future, if Docker fixes the issues)

Mark Wood [10:36 AM]
I will try to stay out of that and just try to find the point where the problem occurs.

Tim Donohue [10:37 AM]
Ok, I don't have anything else to note about Docker stuff today.
Are there any other topics of discussion or notes to add for today? Otherwise, it seems like we might want to close up early (seems to just be you and I, @mwood).

Mark Wood [10:39 AM]
I just wanted to stir up an older issue: is anybody else interested in being able to, for example, substitute DSpace configuration properties in email bodies? DS-3872 uses Velocity instead of, I forget exactly but it's built into the JRE, to do substitutions and provides a way to add more sources of values.

DSpaceSlackBot (IRC) APP [10:40 AM]
qwebirc77345 has quit the IRC channel

Mark Wood [10:40 AM]
(For the longest time I've wanted to remove the need to edit these templates just to change boilerplate to the value of dspace.name, for example.)

Tim Donohue [10:42 AM]
@mwood: I think that's of interest, yes...but, it's higher priority (in my mind) to find an easy way to *translate* those emails. I know the two are not directly related though. But, I hear more complaints about translations than about embedding configurations in email templates

Mark Wood [10:43 AM]
Ah: to have them internationalized/localized. Yes, that should be addressed.

Tim Donohue [10:43 AM]
correct.
So, I'm not against looking at a more powerful translation engine (like Velocity). But, I'd want to ensure it also helps us better support internationalization / localization, as I think that's higher priority than simply embedding configurations easily

Mark Wood [10:45 AM]
I don't recall that EPerson has a "preferred language" so L10N would be controlled by the site admin, for all users?
Perhaps there is a way to get the user's language preference from the browser. I haven't looked at that in a long time.

Tim Donohue [10:47 AM]
EPerson does have a language, actually: https://github.com/DSpace/DSpace/blob/master/dspace-api/src/main/java/org/dspace/eperson/EPerson.java#L188
It's a metadata field these days... "eperson.language"

Mark Wood [10:48 AM]
Ah, that's what I get for trusting my memory. Thanks.

Tim Donohue [10:48 AM]
My main point is that we want to select a templating language based *more* on i18n / l10n than on the ability to embed Configurations. If we find a way to do both easily, that's the best case obviously.
I don't have much more to say about that though. It sounds like a good idea, but may need more analysis.

Mark Wood [10:49 AM]
I think those may be orthogonal, but I agree that L10N of emails is important.
I will look for a Jira issue about this, and if need be, write one.

Tim Donohue [10:49 AM]
I also don't think it's the highest priority right now for DSpace 7.

Mark Wood [10:49 AM]
Also true.
My main point was that this has been sitting around a while and I wanted to see if it had any traction.

Tim Donohue [10:52 AM]
understood. Yes, I think it's a good idea. Just want to keep i18n / l10n in mind as well. Priority is not very high at this time, as there's a lot of other stuff to do for DSpace 7 first.

Mark Wood [10:52 AM]
Right. I just have a list of these that I bring up every so often, when there's time.

Tim Donohue [10:53 AM]
Sounds good
Ok, I have a meeting at the top of the hour (DSpace Steering). Any other final topics / questions / thoughts (have about 6 mins remaining)?

Mark Wood [10:54 AM]
Nothing from me.

Tim Donohue [10:54 AM]
Ok, sounds like it's time to wrap up then! Thanks for the discussion, @mwood!

Mark Wood [10:54 AM]
Thanks, all!

DSpaceSlackBot (IRC) APP [10:56 AM]
*misilot_* has quit the IRC channel

DSpaceSlackBot (IRC) APP [11:02 AM]
misilot has quit the IRC channel

DSpaceSlackBot (IRC) APP [11:23 AM]
qwebirc97115 has quit the IRC channel

Mark Wood [12:48 PM]
Email templates are filled in currently using java.text.MessageFormat.

Mark Wood [1:01 PM]
I created https://jira.duraspace.org/browse/DS-4193 and linked it to all of the previous work I could find on localizing email templates.