DevMtg 2019-03-27

Developers Meeting on Weds, March 27, 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 28 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))
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

    a. **Auto-reindexing in Solr**:

        Unable to locate Jira server for this macro. It may be due to Application Link configuration.

        i. Should this only h
            need to reindex?

    b.   e `authority` core.

        Unable to locate Jira server for this macro. It may be due to Application

        Link configuration.   Or should we use s

4. DSpace (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. PR Build options: https://github.com/DSpace/DSpace/pull/2385
        i. Option 1 - Solve this in the docker build?
        ii. Option 2 - create feature branch (ie configurable entities) when needed
    b. Service initialization and docker integration test script

        i. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/104
    c. Refine Dockerfiles for One Webapp
        i. Keep only spring-rest and rest webapps
        ii. Optional deployment to ROOT
        iii. Feedback for one webapp PR (2265)
    d. 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
    e. ~~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

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

key	summary	type	created	updated	assignee	reporter	priority

Meeting Notes

Meeting Transcript

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

```
Tim Donohue [3:00 PM]
@here: It's DSpace DevMtg time.  The agenda can be found at: https://wiki.duraspace.org/display/DSPACE
/DevMtg+2019-03-27
Let's do a quick roll call to see who is able to join today's discussions


Terry Brady [3:01 PM]
here

Tim Donohue [3:03 PM]
well, it might just be you and I then, @terrywbrady.  I guess we could always do a quick touch-base, and see if
anyone else shows

James Creel [3:03 PM]
Hey there

Tim Donohue [3:03 PM]
Hi @jcreel256.  Welcome

Terry Brady [3:03 PM]
Hello

Tim Donohue [3:04 PM]
Ok, in any case, let's get started and we'll see where discussions take us today
On the DSpace 7 side, I don't have any major updates.  Just a reminder that Configurable Entities feature PRs
need some reviews/eyes :
• https://github.com/DSpace/Rest7Contract/pull/57
• https://github.com/DSpace/DSpace/pull/2376
• https://github.com/DSpace/dspace-angular/pull/372
I'll also note that the DSpace 7 "Preview" release is still (seemingly) on track for an early April release --
assuming Configurable Entities approval comes through by then.
That's basically it. Other updates in tomorrow's DSpace 7 meeting
No updates to share on DSpace 6.x (as is the usual message these days)
Any quick questions/comments before moving along?

Terry Brady [3:06 PM]
none here

Tim Donohue [3:07 PM]
Ok.  I know @mwood said he *might* join us today (from home).  Looks like he's offline now though, so we might
want to loop back around to his Solr upgrade topic (to see if he's able to join in a bit)
```

Terry Brady [3:08 PM]
2058 and the ext solr PR for docker are merged

Tim Donohue [3:08 PM]
So, I'm going to jump over that topic for now...we'll come back to it. The others might be relatively quick anyhow
@terrywbrady: good point. Yes, my links there need updating anyhow
I just updated the meeting agenda to remove those things already merged. Sorry about that :wink:

Mark Wood [3:10 PM]
Hi, sorry, my version of Slack was too old and wouldn't start. Got a new one now.

Tim Donohue [3:10 PM]
Oh, and I see @mwood just joined. Perfect timing! Let's do the Solr upgrade updates/discussion first then
So, in #dev today, I know @mwood was talking about the requirement to *reindex all cores* for the DSpace 7 upgrade. That brought back up the fact that DSpace does auto-reindexing of search core (Discovery)
Which brought back up this ticket/discussion about the current behavior of the auto-reindexer: https://jira.duraspace.org/browse/DS-3658
Simply put, currently (as of DSpace 6), the search core is auto-reindexed anytime any changes to the database occur (i.e. if Flyway runs a migration, it triggers custom code in our DatabaseUtils which trigger a background reindex of the search core)
So, there were two questions here... first, the current behavior is a bit controversial (see DS-3658 ticket), should we change it (or make it configurable)?
second, based on the answer there, do we need to require *manual* reindexing for the DSpace 7 upgrade, or do we keep it automated
Is that accurate, @mwood? Anything else to add?

Mark Wood [3:15 PM]
I think that's it. We definitely need exactly one reindex when upgrading to DSpace 7, neither 2 nor 0.

Tim Donohue [3:16 PM]
Right. If automatic reindexing is gonna be triggered, we don't want people *also* doing it manually :wink:

Terry Brady [3:16 PM]
Since folks will need to explicitly migrate solr data as part of the upgrade, perhaps the manual triggering is appropriate for the 7 upgrade

Tim Donohue [3:16 PM]
(and visa versa)

Mark Wood [3:16 PM]
I haven't found a way to upgrade Solr cores in place, when the classes that wrote them are removed, so you start out with empty cores that need to be filled.
@terrywbrady that is what triggered this discussion: I was writing upgrade instructions, and then saw that DSpace will want to reindex itself *if* we had any fresh migrations.

Tim Donohue [3:18 PM]
@terrywbrady: that seems reasonable to do manual triggering for DSpace 7. But, currently (based on current code) that'd do a double reindex of Discovery/Search. We don't yet have a way to disable auto-reindexing easily (hence DS-3658)

Mark Wood [3:18 PM]
Well, we could just tell folks to delete the reindex flag file.

Tim Donohue [3:18 PM]
So, if we want manual reindexing, we need a way to turn off automatic reindexing (or temporarily disable it)
@mwood: true. Yes, that's the way to temporarily disable auto-reindexing. Delete the reindex flag file manually

Terry Brady [3:19 PM]
I presume that would be easy to implement....

Mark Wood [3:20 PM]
Writing this up clearly will be interesting, because the upgrade instructions are for upgrading to any 7.x from any previous version, including 7.(x-1).
If we know for certain that this upgrade will *always* have an automatic reindex, then alternately I could just take that line out of the manual rebuild instructions.
But perhaps we ought to first figure out whether we want to alter the auto-reindex in some way.

Terry Brady [3:22 PM]
Based on my experience and the comment in the tickets, I think we should stop the auto re-indexing ... or only

allow it to occur if a property is set... or only do it if the core is empty

Mark Wood [3:22 PM]
Discussion in the ticket raises a valid point: that a large site will require hours or days to re-index, and one might want manual control of that.
Making it configurable, default off, does seem like a good approach.

Tim Donohue [3:24 PM]
Auto-reindex seems to definitely have had unintended consequences (especially for large sites). The goal was really to simplify the upgrade process for novice / newer users -- one less step (which means less chance for human mistakes) and it's generally "quick" for smaller sites.

Terry Brady [3:24 PM]
In a demo context (like the preview release) that would give us the option to turn it on explicitly.

Tim Donohue [3:24 PM]
But, as stated, for large sites, it's a big issue if auto-reindex runs frequently or unexpectedly

Mark Wood [3:25 PM]
There is already a PR to make this configurable. It defaults to True rather than False, but that's easily remedied.

Tim Donohue [3:25 PM]
At the very least it does seem like it should be configurable. I'm less certain myself whether it should be "on" or "off" by default.

Terry Brady [3:26 PM]
I suppose a large installation could choose to turn that on...

Tim Donohue [3:26 PM]
personally, I have a tendency to keep the defaults set at "sane values for sites getting started" (which implies it should be "on"). but I see both sides
In any case, yes, the PR seems like something to move forward here
This PR that is: https://github.com/DSpace/DSpace/pull/2184

Mark Wood [3:27 PM]
OK, you convinced me: True is the right default, and we can add instructions on when and why you might want to set it False. (edited)

Terry Brady [3:27 PM]
As you can see, I have advocated both sides in the ticket, the PR, and in this conversation

Tim Donohue [3:28 PM]
Ok, so PR#2184 leaves it set to "true" (or "on") by default. So, it sounds like we're OK with that. We're just giving the power to choose to turn it off

Mark Wood [3:28 PM]
So now I can write something about the 1-time index upgrade.

Tim Donohue [3:29 PM]
And it sounds like we should move PR#2184 along here... get it merged soon (it's tiny enough) so that it can also be worked into the upgrade notes as needed

Mark Wood [3:29 PM]
Yes.

Tim Donohue [3:31 PM]
hmm... just realizing PR#2184 is adding this to 6.4. That's probably OK, and it would *require* leaving it set to "true" (as otherwise we'd change behavior between 6.3 and 6.4).
We'd need to have a corresponding PR or cherry-pick to `master`

Mark Wood [3:31 PM]
I can do that.

Tim Donohue [3:32 PM]
I'm going to make a minor enhancement to the comments in PR#2184 and then give it a :+1: . It looks fine to me, I just want to describe more (in comments) *when* auto reindexing occurs.

Mark Wood [3:32 PM]
I made a note to do that. After it is in, I can update the doco.

Tim Donohue [3:33 PM]
Sounds like a path forward. Anything else to note/add here? Or shall we move on?

Mark Wood [3:34 PM]
I have no more.

Tim Donohue [3:35 PM]
@mwood: anything to say about the dump/restore for Authority core? https://jira.duraspace.org/browse/DS-4187

Mark Wood [3:36 PM]
I have patches ready for 5_x, 6_x, and master. But then I happened to learn that solr-statistics-export/import
also does authority. So, do we want mine?

Tim Donohue [3:36 PM]
Oh, it does?

Mark Wood [3:36 PM]
I have not tested the other one. I only learned of it around noon today.
That's what it claims.

Terry Brady [3:36 PM]
I did not realize it worked on authority

Tim Donohue [3:37 PM]
If we already have it, we should use what we have. But, I'd recommend keeping your work around until we verify
whether it actually works

Terry Brady [3:37 PM]
I enhanced those tools recently and I did not test for that case.

Mark Wood [3:38 PM]
If anyone is desperate for code to read, I would appreciate that person's opinion of mine anyway.
Meanwhile I'll see how Andrea's works, and probably then rework the upgrade instructions. It should not be a
lot of work.
Probably we should rename those commands to solr-export and solr-import, since they do more than statistics.

Tim Donohue [3:40 PM]
Yes, agreed. They should be renamed if they work for any Solr core

Terry Brady [3:40 PM]
I think they write to directories containing the name "statistics".
such as "statistics_export"

Mark Wood [3:41 PM]
If so, that should be fixed too. I'll look into it.

Tim Donohue [3:43 PM]
Ok, sounds good. If we can get this all into one script, that sounds great to me. If this proves difficult,
or it just doesn't work, the two scripts approach is also perfectly OK.

Mark Wood [3:43 PM]
Got it.

Tim Donohue [3:44 PM]
The nice thing here is that...if the current script already works, maybe we don't even need to release a
separate tool to help you dump & migrate your authority core. I know that's a big "if" still though

Mark Wood [3:45 PM]
That would be a nice bonus. I still don't have a stand-alone form.

Tim Donohue [3:45 PM]
Well, let us know what you find, and we'll take it from there!

Mark Wood [3:46 PM]
Will do.

Tim Donohue [3:46 PM]
Ok, sounds like we've wrapped that up. Let's move along (only 15mins left)
On the "One Webapp" topic, my PR is still open for reviews: https://github.com/DSpace/DSpace/pull/2265
@terrywbrady has been doing some early testing (thanks!) and he and I discussed today some of the issues he hit
with running this in Docker

Terry Brady [3:47 PM]
I am working up a PR to align the dockerfiles with these changes https://github.com/tdonohue/DSpace/pull/9
Once I am sure that is working, I will do more testing.

Tim Donohue [3:49 PM]
In any case, you'll see @terrywbrady's most recent review notes issues in Docker. It's possible some (maybe
all) are configuration issues. So, other testing/reviews (especially even not in Docker) are welcome. I think
this all should be working right (and it does have ITs). But, obviously, best to find any major bugs now!
That's the only update I have though. Nothing more to say.

Terry Brady [3:50 PM]
This PR will also officially remove the link webapps/solr from the image

Tim Donohue [3:50 PM]
(Oh, and you'll notice @terrywbrady's PR is to my repo. That's cause it's a PR to my PR#2265) (edited)
Ok, I think that's all there is to say though. So, let's move along for today
Next up is any other Docker updates from @terrywbrady? Other work to share here?

Terry Brady [3:52 PM]
I would like to have a mechanism to build "feature" images for priority PR's.
I created a PR that could accomplish this with a dockerhub build.
https://github.com/DSpace/DSpace/pull/2385
But I do not necessarily love this approach.
Another option would be for us to build "feature" branches in DSpace/DSpace as needed to illustrate changes
that need user input.
If that was done, we could simply auto-build an image each time a feature branch is updated.
Do you all have thoughts on this?

Tim Donohue [3:56 PM]
I guess I'd first ask about the use case here. If the primary goal is to make testing larger PRs *easier*,
then moving towards a centralized "feature" branch approach (like with `configurable_entities`) makes a lot of
sense. Large efforts likely should be central branches anyhow (and they are not always).
If the goal is to make testing *all* PRs easier, a centralized "feature" branch approach won't work so well...
as only Committers can create a centralized "feature" branch...and most PRs don't come from Committers

Terry Brady [3:57 PM]
I think the feature branch approach is cleaner.

Tim Donohue [3:57 PM]
I agree it sounds cleaner

Terry Brady [3:57 PM]
Perhaps we establish a naming convention like feature_...
I will close PR 2385 and add this discussion.

Tim Donohue [3:58 PM]
I'm just pointing out that would mean that we'd only have these Docker builds for *some* PRs...those that have
feature branches

Terry Brady [3:59 PM]
We could revive my PR 2308 (I did a strikethrough on it in the notes) if we want every PR built as an image

Tim Donohue [3:59 PM]
A naming convention for feature branches sounds good to me too

Terry Brady [3:59 PM]
How would you feel about creating such a branch for one_webapp?
I'll make a pitch that I have 2 simple PR ports looking for a review. 2361 and 2362

Tim Donohue [4:01 PM]
If making `one_webapp` into a central feature branch helps with review/testing, I'm OK with that. However, it
will require recreating that PR (obviously).

Terry Brady [4:01 PM]
One last note... this PR (for Docker) contains a test script that verifies that all expected services are
running on localhost within Docker. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/104

Tim Donohue [4:02 PM]
(I probably should have started with a central feature branch anyways...I just didn't realize how many changes
/tweaks would be needed to get it working right)

Terry Brady [4:02 PM]
I have not actually thought through the mechanism for creating a feature branch.... if it disrupts ongoing
reviews we could hold off

Tim Donohue [4:03 PM]
@terrywbrady: I'd say hold off then, until it's figured out. I'd only want to recreate this PR *once* (if it
needs to be done). I'd rather not recreate it now, only to find I need to recreate it again (to better align
with any plans here).
We also could use the *follow-up* PR to "one webapp" as a test scenario. That could be a feature branch, once
"one webapp" is merged/tested (if that happens soon-ish)
Ok, I'm realizing we are over time here...by about 5 mins :wink: I'm not calling any more topics, but if
there's anything else needed to "close up" this discussion, I can hang around a few more minutes

Terry Brady [4:05 PM]
Have a good week everyone!

Tim Donohue [4:05 PM]
You too. Let's consider the meeting closed then. Other discussion can move to #dev or PRs/tickets. Thanks
all!

DSpaceSlackBot (IRC) APP [4:06 PM]
tdonohue has left the IRC channel

Mark Wood [4:06 PM]
Thanks, all!