DevMtg 2019-07-24

Developers Meeting on Weds, July 24, 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, July 25 at 14:00 UTC.
- DSpace 7 Entities Working Group (2018-19): Next meeting is Tues, July 30 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. Ongoing Work
   a. Upgrading Solr Server for DSpace
      i. **Solr**:   Unable to locate Jira server for this macro. It may be due to Application Link configuration.

         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   ls (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.
3. 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.   Unable to locate Jira server for this macro. It may be due to Application Link configuration.
   b.   nection 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.
   c. 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

d. 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

Meeting Notes

Meeting Transcript

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

```
Tim Donohue 10:00 AM
@here: it's time for our general DSpace DevMtg.  Agenda for today is at https://wiki.duraspace.org/display
/DSPACE/DevMtg+2019-07-24

Let's do a quick roll call to see who is joining discussions today
Mark Wood 10:01 AM
Hi
Terry Brady 10:01 AM
hello
Tim Donohue 10:02 AM
Welcome Mark & Terry.  Looks like it may just be the 3 of us today...so, it might be a quick meeting
In any case, we'll get started.
On the DSpace 7 front, I don't have any significant updates to speak of.  But, I'll note what discussions have
gone on in recent working group meetings (in case others lurking are interested)
In the DSpace 7 WG, last week we talked a bit about a few performance issues we've hit (mostly in the REST API
layer).  We already have an idea for how to fix those though via implementing Projection (i.e. only requesting
data from the REST API which is needed for the UI).  Detailed meeting notes at https://wiki.duraspace.org
/display/DSPACE/2019-07-18+DSpace+7+Working+Group+Meeting
Pascal Becker 10:05 AM
hi
:clap:
2

Tim Donohue 10:05 AM
In the DSpace 7 Entities WG, discussion was mostly around OpenAIRE v4 implementation, and some early mockups of
how to create relationships between entities in Submission UI. Notes from that meeting are here: https://wiki.
duraspace.org/display/DSPACE/2019-07-23+DSpace+7+Entities+WG+Meeting
That's basically it for the DSpace 7 updates to share for the week. But, if anyone has additional comments
/questions, I'll pause for a moment
Ok, hearing no questions :slightly_smiling_face:
No updates this week on DSpace 6.x (as usual).  Still in a "waiting for a coordinator" state
That's it for quick updates on my end, so we can move along to updates on other ongoing work
Any updates you want to share @mwood on Solr upgrade for DSpace 7? https://wiki.duraspace.org/display/DSPACE
/Upgrading+Solr+Server+for+DSpace
Mark Wood 10:08 AM
Today I am trying to clean up and restart a test of upgrading a 6.3 instance to 7.0, to test the Solr upgrade
instructions.
:clap:
3

Tim Donohue 10:09 AM
yay!  Sounds great
Assuming all goes well, is the next step to have others try things out? Is there anything left to do in the
outstanding tickets that is required for the upgrade (DS-3658 and DS-4187 linked in agenda)
Mark Wood 10:11 AM
If the test works, that should be all that is needed for upgrading when the statistics core is not sharded.
Figuring out what to do about sharding is Phase 3.
Tim Donohue 10:12 AM
Sounds good.  I was struggling to recall what was in "Phase 3" but now I remember :slightly_smiling_face:
Mark Wood 10:12 AM
The dump/restore tool may not be needed.  I found that the existing tool is more capable than I thought.  I
will test first using that, not the DS-4187 work.
Tim Donohue 10:13 AM
that makes sense. Yes, if we find we can simply close DS-4187 as "won't fix" that's fine by me.
```

And, I'm glad you've had more experience with the existing tool...it seems like we may have simply had a lack of docs there.
In any case, thanks again for the efforts here. Let us know if/when you need additional support or testers.
Sounds like things are progressing nicely though
Pascal Becker 10:14 AM
The tool to dump and reload the statistics core works fine for the authority core too.
:+1:
1

We just tested it for a client.
Mark Wood 10:15 AM
Yes, I had not realized that until I looked at the code.
Pascal Becker 10:15 AM
I'll add a comment.
Tim Donohue 10:15 AM
That fact needs to make it into some docs. Perhaps we should rename that tool eventually too :slightly_smiling_face:
Pascal Becker 10:15 AM
maybe we should document it and maybe give it another name in dspace 7.
Mark Wood 10:16 AM
I'll be happy to work on the documentation if nobody else gets it done first.
:+1:
2

Tim Donohue 10:17 AM
:+1: I think better docs need to be required for Dspace 7. Giving it another name seems like a good idea too, but not required (though if someone wants to take it on, hopefully it wouldn't prove difficult)
In any case, thanks for the updates here. Sounds like we are wrapping up this topic, so I'll move along
Next up, @terrywbrady any updates you'd like to share on DSpace + Docker efforts? https://wiki.duraspace.org/display/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals
Terry Brady 10:19 AM
Sure, would you like to start with a summary of our meeting from yesterday?
Pascal Becker 10:20 AM
https://jira.duraspace.org/browse/DS-2372
Tim Donohue 10:20 AM
Sure, I can start with a summary of the Docker meeting yesterday
Terry Brady 10:20 AM
I'll go through items based on the Goal Nums on the wiki page. Goal 3: http://bit.ly/dspace-launcher-dashboard

terrywbrady.github.io
DSpace Launcher Dashboard
Deploy a DSpace PR in Docker on AWS for Testing
Tim Donohue 10:21 AM
Yesterday, @terrywbrady and I met with some other Tech folks within LYRASIS/DuraSpace representing the Fedora, VIVO and DuraCloud projects. Terry gave us an overview of the DSpace Launcher Dashboard: http://bit.ly/dspace-launcher-dashboard
We had a nice discussion on how this tool works behind the scenes and how it might be applicable to other projects beyond DSpace
The feedback I heard in the meeting & afterwards was all positive (nice work, and lots of cool tools within this dashboard), but the major outstanding question is "would people use this?"
We didn't come to any way to really resolve that question though except for possibly putting it in front of people...and we didn't finalize any plans for doing so yet
So, honestly, it's interesting work...we'll keep discussing it within LYRASIS, but I'm not yet certain of the next steps or how to "prove out" whether it'd get much usage, etc.
Pascal Becker 10:24 AM
I love the tool. I had a similar idea for a long time, but never the time to investigate on it. If people would use it? That is something we would need to try to find out.
Terry Brady 10:24 AM
That seems to be the big question to me as well. Who is our most likely audience to use a tool like this? Repository Managers? Repository Managers who attend DCAT? Repository Managers in other parts of the world?
Tim Donohue 10:24 AM
Terry, anything else you want to add from that discussion?
Terry Brady 10:25 AM
Looking at the wiki page, look at Goal 4. At a minimum, I think we should provide a per branch solution for spinning up instances.
Tim Donohue 10:26 AM
Goal 4: https://wiki.duraspace.org/display/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals#DSpaceDockerandCloudDeploymentGoals-Goal4:Managehostedinstancesof DSpaceforeachsupportedbranchofthesystem
Terry Brady 10:26 AM
The users of a PR tool are more difficult to identify.

I have put out a call on Slack to find folks who want to chat about Goals 6 and 7. Pascal and I will have a chat tomorrow to identify how that discussion could move forward.
Mark Wood 10:27 AM
My guess is that, for any given PR, there would be two users: the developer of the PR, and the person who submitted the Item that resulted in the PR.
Tim Donohue 10:27 AM
Yes, I do see a use case here for potentially/dynamically spinning up several demo sites of DSpace based on different branches (a 5.x, 6.x and 7.x demo site, for example). It'd be kinda nice to do our demo sites even more dynamically, and use the Docker tooling here.
:+1:
1

:point_up: that point is related to Terry's Goal #4
So, even if we don't find users of this tool for PR testing...I think we might want to consider keeping some version of this tool around for demo site management, etc.
Terry Brady 10:28 AM
Let's keep this on the agenda for a bit and see if we have any new inspiration. Tim, it would be great if you could represent these ideas at the user meeting in Minneapolis.
Next week, Pascal or I can provide an update on our conversation tomorrow.
Pascal Becker 10:29 AM
I hope it would help us to find more people to test PRs.
Tim Donohue 10:29 AM
@terrywbrady: Sure, I'll see if I can figure out a way to fit this into the North American DSpace User Group meeting in Minneapolis. I wish there was a Docker talk there that could fit this in, but I can perhaps pull it into my DSpace 7 workshop/talk
Pascal Becker 10:29 AM
The code review must be done by at least one committer, but if someone else could say "I tested it, works as described" that often helps a lot.
Terry Brady 10:29 AM
I wish I had the travel approval to be there to present it!
Pascal Becker 10:30 AM
The tool would make it much easier for people to test PRs. And simple PRs could then be tested even by repository manager that do not know how to compile DSpace at all.
Terry Brady 10:31 AM
One interesting part of our conversation yesterday was whether or not there would be value in building every PR as a docker image. There are multiple ways to accomplish that goal if we decided we wanted to do that. But, it could also be a poor use of compute resources for some PR's.
Mark Wood 10:31 AM
There will probably be more interest in testing features than in bug fixes.
Terry Brady 10:32 AM
Pascal, can you imagine your community of repository managers participating in testing if such a tool existed?
Tim Donohue 10:32 AM
Yes, we had talked about how can you decide which PRs are "big enough" (or important enough) to warrant spinning up in Docker/AWS....likely it's a subset of PRs (maybe even a minority)
Pascal Becker 10:33 AM
I thing in the long run. If someone run into a bug and a PR exist, you could tell them "test the solution and add a comment to the PR" While I can imagine that that would work, it would probably take sometime until people would start testing PRs on their own with that tool.
Can we build PRs that have a specific label?
Terry Brady 10:34 AM
We could do something with tags I suspect.
Pascal Becker 10:34 AM
If we could trigger the build by setting a label, that would be nice.
Terry Brady 10:35 AM
Although perhaps a tag does not exist until a PR is merged.
For the preview release, we did some good stuff with feature branching.
Tim Donohue 10:35 AM
And, as was pointed out in our discussion yesterday...technically, if all this Docker tooling exists & you are creating Docker images per important PR, you could simply spin up a PR test locally in Docker very easily as well. So, this tool may not be as necessary unless you have users who are not willing to run Docker locally, but still want to help test PRs.
Pascal Becker 10:35 AM
git tags wouldn't work as the code is not in DSpace/DSpace until it is merged. Would have to be github labels or something comparable.
Mark Wood 10:35 AM
The process that is notified of new PRs could ask Github for the labels, if it doesn't receive them with the notice. (edited)
Pascal Becker 10:37 AM
It is something completely different to tell someone "install docker and pull this image" or "click here, wait five minutes and start testing".
I would really hope that the tool could help us to get people involved into testing, that were not able to do

that now.
Tim Donohue 10:37 AM
I think my overarching question here is...Do repository managers (and similar non-developers) have a desire to help us test during the development process? If so, this is perfect for that. But, I haven't heard that desire yet, outside of the formalized Testathon (which is a rare event)
Terry Brady 10:38 AM
I wish we could figure out how to create that interest because they are the right audience to participate.
Pascal Becker 10:39 AM
You have persons like helix in the old days who are learning DSpace, are interested in it, and want to try out as much as they can. Besides that you have the testathrons. And then you have repository managers that want to help one particular feature or bug fix.
I think each of the three kinds of use cases I listed would be enough to justify such a tool (edited)
Does the tool produces any costs if it is provided but no one is using it?
Terry Brady 10:40 AM
The cost is small (storage) when the service is unused.
Tim Donohue 10:42 AM
I think we all agree here, honestly, we are just coming at it from different directions. This tool provides a different way of participating in active development, and it could be a useful one if folks are interested in that manner of participation (which I'd also hope they are, but honestly haven't seen that interest yet...but maybe this tool can help generate such an interest)
The other cost here is simply management of the tool...so, it's a person/effort cost. I think those are likely relatively minimal to start as well though (assuming Terry agrees that this tool could be spun up "as-is" as a proof of concept)
Terry Brady 10:43 AM
The code is there to be used!
Tim Donohue 10:44 AM
@terrywbrady: I was more asking is this stable enough for folks to use now...or are there outstanding things that you'd say: "oh, we really should fix this bit here before it is too public"
Pascal Becker 10:45 AM
Yes, part of the problem is, that it required technical skills to test PRs. We don't know if people who misses these would be willing to test until we can tell them "here you go, one click, 5 minutes time and you can test it yourself." As of today these people probably never tested anything because they thought they misses the skills for that.
Tim Donohue 10:45 AM
(@terrywbrady: And it sounded like from yesterday's discussion it's mostly stable...but there were some AWS level configs you had to manually configure that need reconfigured if it is moved elsewhere)
Terry Brady 10:46 AM
It is stable enough to use. I would suggest monitoring the use of the "create instance" button and possibly deciding to grant access to that capability rather than keeping it wide open.
Tim Donohue 10:46 AM
"grant access" : Is that something easy to do in this environment? Or are you talking setting up a login /authentication system
Terry Brady 10:46 AM
Yes, it would take a few hours of collaboration to port this to another AWS account. I would be glad to help there.
Mostly in the API Gateway config.
Tim Donohue 10:47 AM
Oh, ok. Remind me how that works... does that require an AWS account/login? Is it IP based, or something else?
Terry Brady 10:48 AM
Possibly adding a login for that web page. AWS has several options for creating user accounts. I think it can also use Google/social ids for verification.
Tim Donohue 10:48 AM
(I'm not a trained AWS person...I know some of it, but not everything by far)
Terry Brady 10:48 AM
Those are incremental improvements that can be added if usage takes off.
Tim Donohue 10:49 AM
Ok, makes sense that they may not need to be done immediately.
Terry Brady 10:50 AM
Pascal, from your leadership role on the project, if you can think of a way to inspire user participation from the leadership end, that would be a good thing for this effort.
Thanks for the time for this topic.
Tim Donohue 10:51 AM
I suspect, if we wanted to try this out, I'd need to get this approved either through DSpace Leadership /Steering (for the minimal amount of $$ it could add to the budget, which we'd need to better estimate), or find similar approval from within LYRASIS (which unfortunately may take a bit longer simply because merger activities are higher precedence right now...but I can start asking around to see)
Or, the third option, is we find someone willing to host this temporarily for a "trial period" with the hope that it becomes "official" (moves over to DSpace budget) if it sees some interest, etc
Pascal Becker 10:53 AM
I'm totally into that tool. The problem ist: it would take time, until we really see efforts. Even if we advertise it, people will need time to test it, and be reminded on it, and test it again, ...

:+1:
1

Terry Brady 10:53 AM
I have seen some questions in the forums about DSpace Docker in production. It will be good to offer some options. Realistically, I think DSpace 8 will be the target for true support. Pascal may convince me otherwise tomorrow.
@tdonohue, from a $$ perspective I would ask for $100/mo and control usage if the budget is ever hit.
Tim Donohue 10:55 AM
@terrywbrady: thanks for that estimate. Good to know the scale we are talking
So, we are heading to the end of this meeting. This has been a good discussion, and it sounds like there's some ideas to possibly move this forward. Though, I don't feel like the path is 100% clear yet : slightly_smiling_face: Still, it's something we can keep promoting/mentioning (and I'll keep doing so too) and see if we find a way to fund a proof of concept to see if folks will use it
Any final thoughts then for today? I don't have a better wrap-up here, but I've enjoyed the discussion
Ok, not hearing anything. So, let's wrap this up for today. We'll keep this on the agenda and keep brainstorming ways to move it forward.
Thanks for the discussion today, all!
Terry Brady 10:59 AM
Have a good week!
Pascal Becker 11:00 AM
Thank you! Bye, bye! :wave:
Mark Wood 11:00 AM
Thanks!
Pascal Becker 11:00 AM
You too!