

DevMtg 2018-10-17

Developers Meeting on Weds, October 17, 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 \(2016-2023\)](#): The [DSpace 7 Community Sprint 3](#) wrapped up.
- [DSpace 7 Entities Working Group \(2018-19\)](#): Next meeting is Oct 23
 - Last week's meeting (including video recording): [2018-10-09 DSpace 7 Entities WG Meeting](#)
- [DSpace Developer Show and Tell Meetings](#): Next meeting is Oct 30 at 15:00 UTC on [Docker Updates](#)

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. Sprint #3 is ongoing, no updates to report.
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.
3. [DSpace \(Pascal-Nicolas Becker\)](#): [Unable to locate Jira server for this macro. It may be due to Application Link configuration.](#)
4. [Terrence W Brady](#): [DSpace and Docker](#)
5. [DCAT Accessibility Discussion \(Terrence W Brady\)](#): [DCAT Meeting October 2018](#)
6. DSpace 6x: Merge Legacy and Current Statistics: <https://github.com/DSpace/DSpace/pull/1810>
7. Follow-up on "DSpace Top GitHub Contributors" site (Tim Donohue): <https://tdonohue.github.io/top-contributors/>
8. Brainstorms / ideas (Any quick updates to report?)
 - a. Follow-up on Curation Task Reporting (PR 2180)
 - b. [Bulk Operations Support Enhancements](#) (from Mark H. Wood)
 - c. [Curation System Needs](#) (from Terrence W Brady)
 - i. PR 2181 implements per-run task parameters. Ready for review.
 - ii. PR 2180 improves reporting. Ready for review.
9. 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)

Alexander Sulfrian [10:00 AM]
Hi

Tim Donohue [10:01 AM]
@here: It's time for our weekly DSpace DevMtg. Agenda is at: <https://wiki.duraspace.org/display/DSPACE/DevMtg+2018-10-17>

Terry Brady [10:01 AM]
hello

Tim Donohue [10:01 AM]
And, yes, let's do a quick roll call to see who is joining today

Mark Wood [10:01 AM]
Hi

Tim Donohue [10:02 AM]
Hi all. I'll go ahead and start with a few quick notes/reminders...we may have a few more folks joining in a bit
Last week was our 3rd DSpace 7 Sprint! You may have seen the summary on mailing lists, but it's also here:
<https://wiki.duraspace.org/display/DSPACE/DSpace+7+Community+Sprint+3>

Terry Brady [10:03 AM]
Congratulations on the Sprint!

Pascal Becker [10:03 AM]
hi

Tim Donohue [10:04 AM]
In this week's DSpace 7 meeting (tomorrow at 14UTC), we'll be setting aside the entire meeting for DSpace 7 Strategic Discussions.... The Steering Group meet earlier this week and wanted some updates on DSpace 7 status, and what we feel we can release in an "early beta" later this year.
Any are welcome to join that discussion, but it will be higher level than normal -- more about upcoming development plans & strategy. Less about individual PRs (unless we have time at the end)
The next DSpace 7 Entities WG meeting is next Tues (Oct 23). See agenda for more.
And the next Dev Show & Tell is a week after that, on Oct 30
I think those are the main quick updates/reminders to add in here

Terry Brady [10:06 AM]
Thanks for the reminder on the Show and Tell. I will make a note to send reminders next week.

Tim Donohue [10:07 AM]
Moving into our discussion topics.. I don't have any other updates on 7.x or 6.x to speak of today...so, I'm going to skip over topics #1 and #2, as I think the main topics are in #3-6 in are agenda.
First up then is Deleting EPersons in DSpace (@pbecker): <https://jira.duraspace.org/browse/DS-4036>
I'm not sure if this needs much introduction, as the goal is pretty self-explanatory. We need to be able to delete (or anonymize) EPerson objects out-of-the-box to meet guidelines like GDPR
While it is possible to currently remove any identifying information from an EPerson object (i.e. anonymize them), it's a very manual process right now, and we'd like to move towards supporting full *deletion* of these objects (like you can delete anything else in DSpace)
@pbecker has an early PR that does that exactly: <https://github.com/DSpace/DSpace/pull/2229> And it is waiting for reviews/feedback

Pascal Becker [10:11 AM]
There were even doubts if you can really remove all identifying information, as you can also see which items a eperson submitted and look if they all match one single author.

Tim Donohue [10:12 AM]
@pbecker: Yes, true , full anonymization may not always be possible...depending on whether EPeople are submitting their own papers or if someone is submitting on their behalf

Pascal Becker [10:13 AM]
There was a discussion about which way to go and that is something we should pick up here again.
The problem we have is that we cannot delete an EPerson as soon as it is referenced anywhere in DSpace. Mostlay Epersons are referenced as item submitter.
I decided to set those references null and to change DSpace in a way that it can handle a null submitter instead of running into a NPE.

Terry Brady [10:14 AM]

Would it be possible for a malicious user to deposit several harmful items into a repository and then delete their account? Is there a need to identify items from the same user for that purpose? I suppose a malicious user could also generate multiple accounts.

Pascal Becker [10:14 AM]

I think this is the cleanest way to do it.

Tim Donohue [10:14 AM]

Personally, I'd like to see full EPerson deletion possible here. I.e. I'd like to move forward with @pbecker's PR#2229 (linked above). But, I'd like to hear what others think on this

Pascal Becker [10:15 AM]

@terrywbrady we still have the information in dc.description.provenance. And I think we may keep it there for exactly the reason you just pointed out.

So while we are deleting personal accounts, we of course keep the information who submitted an item in case of any legal troubles.

Mark Wood [10:15 AM]

The code should be able to handle a null submitter, regardless whether we provide a way to destroy EPersons.

Terry Brady [10:15 AM]

Good to know. Does that provenance info raise any GDPR concerns?

Tim Donohue [10:16 AM]

Could we consider limiting EPerson deletion to Administrators only? Would that be good enough for GDPR (I thought GDPR just needs the ability for individuals to request their acct be deleted -- not the ability for them to delete it *themselves*)

Mark Wood [10:16 AM]

Copy requests are another possible concern. There is no 'delete' operation for these.

Pascal Becker [10:17 AM]

GDPR works like this: you may store information only if you have a good reason to do so. Being able to answer to jurisdictional requests seems to be a good reason for me. But of course I'm neither a lawyer nor a data protection expert.

@tdonohue I think deletion through administrators should be enough.

Mark Wood [10:17 AM]

"good reason" will need to be adjudicated for various claims of "good reason," I think. We won't have definitive answers for some time.

Pascal Becker [10:18 AM]

@mwood Our PR sends the copy requests to administrators or a helpdesk email address if IIRC.

I think we start with deleting EPersons. If we notice later that we have to cleanup dc.description.provenance, we can do that later. But I don't think we will have to do so.

Mark Wood [10:18 AM]

The request is stored in the database to support copy-request workflow.

Tim Donohue [10:19 AM]

Yes, the PR sends copy requests to Admin if the eperson is deleted. I saw that in PR#2229

Pascal Becker [10:19 AM]

Ah, you're talking about requests made, but not answered yet.

Mark Wood [10:19 AM]

AFAICT copy requests are *never* deleted.

Pascal Becker [10:19 AM]

I think we probably just delete those. We could discuss if they must be resend to administrators.

Tim Donohue [10:20 AM]

@mwood: that should be logged as a separate bug then. I wasn't aware copy requests are kept forever

Mark Wood [10:20 AM]

But copy requests are a separate issue from EPerson deletion.

Tim Donohue [10:20 AM]

I'd like to concentrate this discussion on EPerson deletion though

Mark Wood [10:21 AM]

OK, had to think it through a bit.

Tim Donohue [10:21 AM]

I.e. Should we support full deletion of EPerson objects? If so, are there others here willing to help us review / test the PR#2229 (from Pascal)?

DSpaceSlackBot (IRC) APP [10:21 AM]

qwebirc2166 has quit the IRC channel

Tim Donohue [10:22 AM]

Personally, I think <https://github.com/DSpace/DSpace/pull/2229> looks like a great direction, and it comes with Unit Tests to prove out the new logic (after the Eperson is deleted / set to "null"). But, it needs more eyes

Terry Brady [10:22 AM]

I added myself as a reviewer

Pascal Becker [10:23 AM]

I must mention that we ported this PR from DSpace 6 and did not update it fully properly yet. We need help from the DSpace7-Rest- and DSpace7-Angular-Team, but we also must look out for new code that deals with references to EPersons and where added between we touched our PR the last time and it was published.

Tim Donohue [10:23 AM]

It also needs more testing to ensure it covers all the cases/scenarios -- i.e. nothing has been overlooked (as we know, we don't have unit tests for *all* DSpace webapps/functionality)

Pascal Becker [10:23 AM]

I hope I find some more time myself next week. I want to go through the code and take a look again at all places that works with EPersons.

The basic topics for today were to raise attention for this and to ask about objections against the general strategy (setting references null).

Mark Wood [10:24 AM]

Entities WG will need to be sensitive to this issue too, for entities which represent natural persons.

Tim Donohue [10:25 AM]

@mwood: yes, of course. But, I suspect less issues there, as Entities are Items and all Items can already be deleted. So, if there's an Entity which is a natural person, we should be able to delete it like any other item

Mark Wood [10:26 AM]

@pbecker I was uneasy about this approach at first, but on reflection I have no objection.

Pascal Becker [10:26 AM]

@tdonohue @mwood made a good point here: we need to raise awareness for that.

@mwood thank you. :slightly_smiling_face:

Tim Donohue [10:27 AM]

I agree, though I know the Entities WG participation is mostly from Europe...they won't want to build anything that doesn't align with GDPR out-of-the-box :wink: But, point is well taken

Pascal Becker [10:28 AM]

I was astonished when I noticed that a null submitter lets DSpace run in a lot of NPEs. As @mwood said before, DSpace should be able to deal with this, independently from the question of deleting EPersons. (edited)

@tdonohue I think we covered it well. Thanks for taking it on the agenda.

Tim Donohue [10:29 AM]

So, it sounds like we all approve/agree on this direction, which is good!

I'll also sign up to review this PR. The more eyes the better, so please others take a look / give it a test.

This is very high priority for 7.x, and we want to get it right

So, that wraps up that topic...moving along

Next up was an update on Docker PRs from @terrywbrady. <https://wiki.duraspace.org/display/DSPACE/DSpace+and+Docker#DSpaceandDocker-Proposal:SupportMultipleJDKVersions>

Terry Brady [10:31 AM]

I have a master branch PR and ports for 6x, 5x, and 4x. It would be very useful to merge this before the Show and Tell.

@pbecker much of this work was done based on requests/suggestions from you.

Pascal Becker [10:33 AM]

I like it very much!

Terry Brady [10:34 AM]
Could I get some +1's?
@tdonohue has reviewed the main PR.

Tim Donohue [10:34 AM]
Here's the PRs needing review:

- <https://github.com/DSpace/DSpace/pull/2214>
- <https://github.com/DSpace/DSpace/pull/2217>
- <https://github.com/DSpace/DSpace/pull/2218>
- <https://github.com/DSpace/DSpace/pull/2219>

Terry Brady [10:34 AM]
Thanks for the list

Tim Donohue [10:35 AM]
I can re-review the main PR (#2214) based on my previous feedback. It was close before, I only had some minor suggestions which look to be implemented now

Terry Brady [10:35 AM]
Built images are here (until the merge): <https://hub.docker.com/r/terrywbrady/dspace/builds/>

Tim Donohue [10:36 AM]
@pbecker: Are you able to give these PRs a quick review too?

Pascal Becker [10:36 AM]
Already on it.
I just see the note that tomcat8 and DSpace 5 is not recommended to be used together. What is the reason?

Tim Donohue [10:37 AM]
DSpace 5 installation instructions say Tomcat 7.

Terry Brady [10:37 AM]
The classpath resolution is different. additions.jar is not found by tomcat8
In tomcat7, it is found by listing jars alphabetically.

Tim Donohue [10:38 AM]
While you *can* run DSpace 5 with Tomcat 8 in some scenarios, @terrywbrady is right, you need to make tweaks if you want to use `additions.jar` concept
So, we were trying to keep this simple...and just leave DSpace 5 with Tomcat 7

Pascal Becker [10:38 AM]
I see. And in DSpace 6 we solved this with bumping the servlet version and changing the web.xmls?

Terry Brady [10:38 AM]
@tom_desair has a documented resolution for the issue, but we decided to have the images match the installation instructions.

Mark Wood [10:38 AM]
The ordering of the classpath was never specified in the Servlet spec. Tomcat 7 sorts it; Tomcat 8+ do not.

Tim Donohue [10:39 AM]
@pbecker: correct. We solved this bug in DSpace 6. It wasn't backported to 5.x, as 5.x was released for Tomcat 7.
Here's the specific bug that was fixed in 6.x: <https://jira.duraspace.org/browse/DS-2437>

Pascal Becker [10:40 AM]
:+1:

Tim Donohue [10:42 AM]
Ok, in any case, it sounds like we have reviewers for these Docker PRs. As noted, it'd be good to get them reviewed/merged prior to the next Dev Show & Tell (Oct 30)
Moving along then to our next topic... DCAT Accessibility Discussion (also @terrywbrady) : <https://wiki.duraspace.org/display/cmtygp/DCAT+Meeting+October+2018>

Terry Brady [10:43 AM]
We had a really good discussion about accessibility and the complexity of defining requirements.
Unfortunately, there was not a consensus to create the accessibility working group that @tdonohue proposed.
I am worried that we are going to release DSpace 7 and still leave every implementing institution on their own to figure out their requirements.
I recommend escalating this issue (to the steering group?) to see if a working group could be established.

Mark Wood [10:46 AM]

A good deal of accessibility is always going to be on the individual site. If they alter themes to use unfortunate color combinations, for example.

Pascal Becker [10:46 AM]

What where the arguments against a WG? Or was it just not possible to find people willing to join a WG?

Tim Donohue [10:47 AM]

Is there an opportunity to at least get passionate folks/institutions to be more "proactive" about this? We have a DSpace 7 demo site which can be tested for accessibility *now*. If we are already doing things wrong, I'd like to see bug tickets logged, so that I can ensure they get fixed for 7.0

Terry Brady [10:47 AM]

I think it was a lack of initiative and time rather than thinking the idea was a bad one.

Tim Donohue [10:48 AM]

I'll be honest here, while I can bring this up with Steering & DSpace 7 team...the initial response will very likely be "we need to find volunteers to help with this"

Terry Brady [10:48 AM]

There are some high level issues to think about such as "is alt-text noise or is it useful? if it is useful, what should go into the tag"

Pascal Becker [10:48 AM]

@cjuergen was always involved in accessibility. Maybe we could ask her to test DSpace 7 in regard of accessibility?

Tim Donohue [10:49 AM]

If there is a lack of space to discuss these topics, we also could do this more informally...create an #accessibility channel in Slack, and encourage folks to join if they have interest -- and keep pestering folks to help us test DSpace 7 :wink:

I think this is definitely an area that we need external testers/feedback. While the DSpace 7 team can be tasked with fixing accessibility bugs, finding external testers would be ideal

Terry Brady [10:51 AM]

It seems like this could be "solved" once. Create a set of recommended approaches and document the rationale. Otherwise, requirements become very subjective.

Tim Donohue [10:53 AM]

So, if someone (even Georgetown?) could help us with even a proposal, that might help move this forward. There are *ways* to do some amount of automated accessibility testing in Angular via Travis CI...see: <https://github.com/DSpace/dspace-angular/issues/209>

But, as you said, we need guidelines/recommendations to strive towards, and then we'd need someone(s) to help select a tool to test with.

I'll admit, it's just a tad frustrating to hear DCAT say "In DSpace 7, accessibility is very important" but not be willing/able to help us figure out what we should strive towards.

Mark Wood [10:55 AM]

I know in a general way that there are tools aimed at US CFR 9 ss504/508 compliance. I don't even know what to ask about in EU and elsewhere.

Terry Brady [10:56 AM]

It was a really useful conversation. We shared our approach to alt-text on thumbnails. Kimberly from U of AZ shared feedback they had received on alt-text. The feedback was basically that the alt-text could be perceived as noise.

Tim Donohue [10:56 AM]

So far, the feedback on <https://github.com/DSpace/dspace-angular/issues/209> has been for "WCAG 2.0 AA" compliance.

Terry Brady [10:57 AM]

A really constructive conversation, but the solutions could go in very different directions.

Mark Wood [10:58 AM]

The infrastructure used for DSpace 7 already tries to get aspects of accessibility right. Could our initial strategy be simply, "try not to break that"? We're not going to become perfectly accessible in a single release. How close might that get us?

Tim Donohue [11:00 AM]

@mwood: yes, though I've been hoping we could get an external expert to do some testing to see how we are doing

so far. I agree though, Angular has a lot of built in accessibility that "just works"

Terry Brady [11:02 AM]

My hesitation here is that I think it may be premature to nail down a technical solution. I think we need our librarians and administrators to define what accessibility means for our systems. Once we know that, I think we can find technical solutions.

And our updated frameworks in DSpace 7 should make those solutions much easier.

Tim Donohue [11:03 AM]

Understood, @terrywbrady. I can agree with that. I think I'd still like to see someone (any institution with an accessibility expert/tester on staff) do some initial accessibility tests on DSpace 7. It could happen anytime, though maybe we really push for later this year when the "beta" comes out

Terry Brady [11:04 AM]

Thanks for the time!

Pascal Becker [11:04 AM]

@tdonohue we really should ask @cjuergen if she can help out.

Tim Donohue [11:04 AM]

Sure thing, and thanks for the feedback! I'm really glad DCAT is discussing this, I just hope maybe we'll find some interested folks / institutions to help us move this idea forward

@pbecker: she should already get notified by your mentions :wink: Hopefully she's interested

As we're now over time here...I think we'll unfortunately need to leave all other topics for next week

Mark Wood [11:05 AM]

If anyone is interested, I have a browser extension "WAVE" installed, which catches things like missing alt text, broken ARIA references, etc. I don't yet know how to use it.

Terry Brady [11:06 AM]

This is an area where I think DCAT is the right group to champion the issue. Unfortunately, the nature of that group makes it difficult for DCAT to take ownership of issues.

Tim Donohue [11:06 AM]

But, before we wrap up, I do know that @terrywbrady wants more feedback on whether to push <https://github.com/DSpace/DSpace/pull/1810> into 6.x (i.e. move it from 7.0 to a 6.4). We can push that to the top of the agenda for next week, but feedback also welcome in GitHub

Unfortunately, we'll have to wrap up our discussion for today. I'll draft up an agenda for next week. Thanks all for the discussion today!

Terry Brady [11:07 AM]

Thanks