

DevMtg 2019-06-26

Developers Meeting on Weds, June 26, 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](#): Next meeting is Thurs, June 27 at 14:00 UTC. Agenda: [2019-06-27 DSpace 7 Working Group Meeting](#)
- [DSpace 7 Entities Working Group](#): Next meeting is Tues, July 2 at 15:00 UTC. Agenda: [2019-07-02 DSpace 7 Entities WG Meeting](#)
- [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. (Ongoing Topic) [DSpace 7 Status Updates](#) for this week (from [DSpace 7 Working Group](#))
 - a. [DSpace 7 Preview Release](#) is out the door!
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. **Auto-reindexing in Solr:**

↑ [DS-3658](#) - Full reindex (index-discovery) is triggered whenever Flyway finds a new DB migration CLOSED

 - i. Should this only happen for **major** releases? Should it be configurable? Can we find a more precise trigger? When **do** we need to reindex?
 - b. Dump/restore tool for the `authority` core.

+ [DS-4187](#) - Tool to dump and restore the 'authority' Solr core CLOSED

 Or should we use `solr-export-statistics`?
4. [DSpace Docker and Cloud Deployment Goals \(old\)](#) (Terrence W Brady)
 - a. 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
 - b. [DSpace Launcher Dashboard - Deploy a PR on AWS for Testing](#)
 - i. There is a 2 minute video that illustrates this proposal.
5. 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>
 - b. [DSpace Backend as One Webapp](#) - First phase is complete, but second phase will involve renaming webapp to be "dspace-server"

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. 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	T	Created	Updated	Assignee	Reporter	P	Status	Fix Version /s
DS-3962	Improve community-list performance		Jul 23, 2018	Jul 10, 2020	Unassigned	April Herron (Atmire)		CODE REVIEW NEEDED	6.5
DS-3952	Expose Request Copy		Jul 12, 2018	Aug 03, 2021	Mark H. Wood	Mark H. Wood		CODE REVIEW NEEDED	7.1
DS-4303	Old (v4-6) REST API is deprecated in v7, remove in v8		Jul 17, 2019	Mar 02, 2020	Unassigned	Tim Donohue		CODE REVIEW NEEDED	8.0
DS-4580	Authority control doesn't update when editing item		Apr 25, 2021	Apr 25, 2021	Hrafn Malmquist	Hrafn Malmquist		CODE REVIEW NEEDED	6.4
DS-3849	REST API items resource returns items in non-deterministic order		Feb 23, 2018	Mar 28, 2020	Bram Luyten (Atmire)	Andrew Bennet		CODE REVIEW NEEDED	6.4
DS-4562	Database Connection Pooling: change maxWait default value		Dec 08, 2020	Dec 08, 2020	Unassigned	Hrafn Malmquist		CODE REVIEW NEEDED	6.4
DS-4291	Browse Starts-With 0-9 only returns results for 0		Jul 03, 2019	Mar 28, 2020	Unassigned	April Herron (Atmire)		CODE REVIEW NEEDED	6.4
DS-3453	OAI-PMH verb Identify doesn't work on Oracle		Jan 13, 2017	Mar 28, 2020	Unassigned	Anonymous (No Reply)		CODE REVIEW NEEDED	6.4
DS-4157	Save and Exit on workflow edit metadata		Feb 01, 2019	Mar 12, 2019	Unassigned	Philip Vissenaekens (Atmire)		CODE REVIEW NEEDED	6.4
DS-4201	Starts_with parameter shouldn't be passed when browsing an item index		Mar 28, 2019	Jun 24, 2019	Hrafn Malmquist	Hrafn Malmquist		CODE REVIEW NEEDED	6.4
DS-3996	The "first few letters" search input in the browse (authors, subject, etc) works poorly		Aug 29, 2018	Oct 29, 2020	Unassigned	Santiago Tettamanti (SEDIC)		CODE REVIEW NEEDED	6.4
DS-4190	Word-break CSS class breaks words without hyphens in Firefox and Opera mini		Mar 15, 2019	Mar 15, 2019	Hrafn Malmquist	Hrafn Malmquist		CODE REVIEW NEEDED	6.4
DS-2852	Discovery search labels show authority key instead of author name		Oct 29, 2015	May 18, 2020	Unassigned	Bram Luyten (Atmire)		CODE REVIEW NEEDED	6.4
DS-4382	BTE ignores fields on import if they are configured in input-types.xml using value pairs		Nov 11, 2019	Nov 11, 2019	Pascal-Nicolas Becker	Pascal-Nicolas Becker		CODE REVIEW NEEDED	6.4

DS-4260	CLI harvest promotes non-existing command		May 25, 2019	Aug 03, 2021	Pascal-Nicolas Becker	Pascal-Nicolas Becker		CODE REVIEW NEEDED	6.4
DS-4044	DataCiteConnector assumes schema version 2.2		Oct 09, 2018	Oct 25, 2018	Mark H. Wood	Mark H. Wood		CODE REVIEW NEEDED	5.11
DS-4472	Backporting - Submission Resource Policy not correctly removed during XMLWorkflow		Apr 02, 2020	Apr 02, 2020	Luigi Andrea Pascarelli (4Science)	Luigi Andrea Pascarelli (4Science)		CODE REVIEW NEEDED	5.11
DS-4439	Orcidv2AuthorityValue loses ORCID info stored in solr record		Feb 21, 2020	Oct 07, 2020	Luigi Andrea Pascarelli (4Science)	Yana De Pauw (Atmire)		CODE REVIEW NEEDED	5.11, 6.4
DS-2715	add ORCID integration for JSPUI		Aug 20, 2015	Jun 15, 2020	Luigi Andrea Pascarelli (4Science)	Gustavo		CODE REVIEW NEEDED	5.11, 6.4
DS-4327	Configurable workflow: Provenance metadata lacking submitter details		Aug 21, 2019	Oct 08, 2019	Unassigned	Bram Luyten (Atmire)		CODE REVIEW NEEDED	5.11, 6.4

Showing 20 out of 23 issues

2. Newly created tickets this week:

Key	Summary	T	Created	Assignee	Reporter	P	Status
DS-4589	Node.js version used in "easy" Mirage 2 build is ancient		Aug 02, 2021	Unassigned	Alan Orth		CODE REVIEW NEEDED

1 issue

3. Old, unresolved tickets with activity this week:

Key	Summary	T	Created	Updated	Assignee	Reporter	P	Status
DS-4537	OAI index doesn't update for deleted items		Jul 14, 2020	Aug 03, 2021	Unassigned	Andrew Wood Atmire		CODE REVIEW NEEDED
DS-4528	Scripts & processes: Integration test for parallel processes		Jun 18, 2020	Aug 03, 2021	Unassigned	Kevin Van de Velde (Atmire)		RECEIVED
DS-4493	ItemImport initialises language code with empty string when absent instead of null		Apr 27, 2020	Aug 03, 2021	Unassigned	Bram Luyten (Atmire)		CODE REVIEW NEEDED
DS-4477	Implement Resource Policy patch operations for policyType and action properties		Apr 07, 2020	Aug 03, 2021	Unassigned	Giuseppe Digilio (4Science)		VOLUNTEER NEEDED
DS-4460	Define an unique way to identify type in the REST API		Mar 16, 2020	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		RECEIVED
DS-4446	Add better checks before moving items between collections and remove items from collections		Feb 28, 2020	Aug 03, 2021	Unassigned	Pascal-Nicolas Becker		RECEIVED
DS-4405	ShibolethAuthentication empty header configuration are not ignored		Dec 19, 2019	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		RECEIVED
DS-4399	pdfbox and encrypted PDFs		Nov 28, 2019	Aug 03, 2021	Unassigned	Ivan Masár		RECEIVED

DS-4397	speed up controlled vocabulary loading		Nov 28, 2019	Aug 03, 2021	Unassigned	Ivan Masár		CODE REVIEW NEEDED
DS-4388	LDAP auth groupmap maps to a random value in case of multi-value attributes		Nov 19, 2019	Aug 03, 2021	Unassigned	Ivan Masár		ACCEPTED / CLAIMED
DS-4245	NPE in CSV export		May 10, 2019	Aug 03, 2021	Unassigned	Ivan Masár		MORE DETAILS NEEDED
DS-4242	Analyze if metadata "place" field can be removed in favor of JSON array ordering in REST API		May 06, 2019	Aug 03, 2021	Unassigned	Tim Donohue		VOLUNTEER NEEDED
DS-4168	Performing metadataschema or metadatafield PUT requests as an unprivileged user results in a 404 instead of a 403		Feb 14, 2019	Aug 03, 2021	Unassigned	Paulo Graça		VOLUNTEER NEEDED
DS-4051	Create an eperson with the same email of an existent user fails with a 500		Oct 11, 2018	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		VOLUNTEER NEEDED
DS-4038	Count method of the DSpaceRestRepository is wrong implemented		Oct 06, 2018	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		VOLUNTEER NEEDED
DS-3826	Use an HandlerMethodArgumentResolver to create the Patch object		Feb 01, 2018	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		VOLUNTEER NEEDED
DS-3778	Document the use of the Range header in the REST Contract		Dec 02, 2017	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		VOLUNTEER NEEDED
DS-3754	Review the SubmissionConfigReader and DCInputsReader to introduce enums		Nov 08, 2017	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		VOLUNTEER NEEDED
DS-3675	EPersons cached as submitters aren't uncached when their item is		Aug 17, 2017	Aug 03, 2021	Unassigned	Pascal-Nicolas Becker		VOLUNTEER NEEDED
DS-3527	Add support for etag in GET requests		Mar 16, 2017	Aug 03, 2021	Unassigned	Andrea Bollini (4Science)		VOLUNTEER NEEDED

Showing 20 out of 22 issues

4. Tickets resolved this week:

Key	Summary	T	Created	Assignee	Reporter	P	Status	Resolution
DS-4575	Wrap around DBCP2 could provide more configurable options		Feb 24, 2021	Unassigned	Hrafn Malmquist		CLOSED	Fixed
DS-4574	Upgrade DBCP dependency		Feb 23, 2021	Unassigned	Hrafn Malmquist		CLOSED	Fixed
DS-4543	Item Import SAF should handle multiple metadata schemas		Aug 06, 2020	Unassigned	Jozsef Marton		CLOSED	Declined
DS-4505	Performance issues with finding Collections/Communities a user has Admin rights on		May 15, 2020	Unassigned	Tim Donohue		CLOSED	Fixed
DS-4440	GDPR - Anonymize statistics feature		Feb 25, 2020	Unassigned	Samuel Cambien (Atmire)		CLOSED	Fixed
DS-4396	Shibboleth authentication REST implementation		Nov 27, 2019	Giuseppe Digilio (4Science)	Giuseppe Digilio (4Science)		CLOSED	Fixed

DS-4364	Editing Item Metadata Breaks Relationships		Oct 16, 2019	Unassigned	April Herron (Atmire)		CLOSED	Duplicate
DS-4257	Deploy Java Backend as a single (Spring Boot) webapp		May 21, 2019	Tim Donohue	Tim Donohue		CLOSED	Fixed
DS-4239	Migrate the workflow.xml to spring		May 02, 2019	Kevin Van de Velde (Atmire)	Kevin Van de Velde (Atmire)		CLOSED	Fixed
DS-4223	Metadata Schemas for configurable entities		Apr 25, 2019	Ben Bosman	Ben Bosman		CLOSED	Fixed
DS-4205	Upgrade to Handle Server v9		Apr 02, 2019	Tim Donohue	Tim Donohue		CLOSED	Fixed
DS-4181	Expose the CollectionRole and Workflow configuration over REST		Mar 06, 2019	Unassigned	Andrea Bollini (4Science)		CLOSED	Fixed
DS-4110	SolrUpgradePre6xStatistics: Exception: Invalid shift value in prefixCoded bytes (is encoded value really an INT?)		Dec 12, 2018	Terrence W Brady	Terrence W Brady		CLOSED	Fixed
DS-4071	webui.bitstream.order.* configurations unused in DSpace 6.x		Oct 31, 2018	Unassigned	Tim Donohue		CLOSED	Fixed
DS-4045	Cleanup the submission steps hierarchy		Oct 09, 2018	Unassigned	Andrea Bollini (4Science)		CLOSED	Fixed
DS-3940	Sherpa Romeo old API sunset & transition to new API		Jul 03, 2018	Kim Shepherd	Bram Luyten (Atmire)		CLOSED	Fixed
DS-3860	Remove any usage of HttpSession		Feb 28, 2018	Unassigned	Tom Desair		CLOSED	Fixed
DS-3825	Improve exception handling in the Submission classes		Feb 01, 2018	Luigi Andrea Pascarelli (4Science)	Andrea Bollini (4Science)		CLOSED	Fixed
DS-3747	Changes to the input-forms.xml		Nov 02, 2017	Unassigned	Andrea Bollini (4Science)		CLOSED	Fixed
DS-3744	Configuration changes in DSpace 7		Nov 01, 2017	Unassigned	Andrea Bollini (4Science)		CLOSED	Fixed

Showing 20 out of 29 issues

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

Key	Summary	T	Created	Updated	Assignee	Reporter	P
DS-4586	Item versioning: enable permanent handle reference to the most recent version of each items		Jul 21, 2021	Jul 22, 2021	Unassigned	Jozsef Marton	
DS-4584	Cross site scripting		Jun 08, 2021	Jun 09, 2021	Unassigned	Jaime Solorzano	
DS-4583	SQL injection		Jun 08, 2021	Jun 09, 2021	Unassigned	Jaime Solorzano	
DS-4576	Export/import of Solr authority core with SolrImportExport not handling ORCID "label" multi-value fields properly		Feb 25, 2021	Feb 25, 2021	Unassigned	Sean Kalynuk	
DS-4573	Filter Media relies on Bitstream Name rather than UUID, skipping files with identical names		Feb 23, 2021	Feb 23, 2021	Unassigned	Claudia Jürgen	
DS-4572	Deleted Bitstream corresponding extracted files not deleted		Feb 23, 2021	Feb 23, 2021	Unassigned	Claudia Jürgen	

DS-4571	Registration allows to register more than once with the same email		Feb 18, 2021	Feb 18, 2021	Unassigned	Santiago Tettamanti (SEDICI)	
DS-4570	UUID fragments included in Legacy Usage Reports (log file based usage statistics)		Feb 12, 2021	Feb 12, 2021	Unassigned	Anne Lawrence	
DS-4569	Date Display in Browse Indices		Feb 01, 2021	Feb 01, 2021	Unassigned	Claudia Jürgen	
DS-4568	Possible to create multiple epersons with the same email address		Jan 29, 2021	Jan 29, 2021	Unassigned	Anne Lawrence	
DS-4567	Solr statistics error		Jan 26, 2021	Jan 26, 2021	Unassigned	Humberto Blanco	
DS-4566	Export search results feature incorrectly escapes query parameters in search filters		Jan 26, 2021	Jan 26, 2021	Unassigned	Alan Orth	
DS-4565	Unable to update archived Item using SWORDv2 as collection / community administrator		Jan 19, 2021	Jan 19, 2021	Unassigned	Jakub íhák	
DS-4558	RTS - Update bagit-support to 1.1.0		Oct 21, 2020	Oct 21, 2020	Michael Ritter	Michael Ritter	
DS-4556	Replication Task Suite: Support the reporting of individual failures when part of a larger curation action		Oct 14, 2020	Oct 14, 2020	Unassigned	Bill Branan	
DS-4553	XMLUI generates useless links		Sep 29, 2020	Sep 29, 2020	Unassigned	Hrafn Malmquist	
DS-4550	OpenAIRE and dc.relation		Sep 01, 2020	Sep 02, 2020	Unassigned	Gabriela Mircea	
DS-4549	Significant performance issues when displaying items with many (100+) bitstreams		Sep 01, 2020	Sep 01, 2020	Unassigned	Martin Walk	
DS-4547	Google Analytics code should be moved to <head>		Aug 25, 2020	Aug 25, 2020	Unassigned	Bill Branan	
DS-4546	No JSPUI index community list when changing language		Aug 25, 2020	Aug 25, 2020	Unassigned	Alex	

Showing 20 out of 92 issues

Meeting Notes

Meeting Transcript

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