2019-10-15 DSpace 7 Entities WG Meeting

- Date & Location
- Attendees
- Agenda
- Current Work
  - Tickets to Resolve
  - PRs Needing Review
  - PRs Merged this week!
- Task List
- Notes

Date & Location
15 Oct 2019 at 15:00 UTC (11:00am EST)

Location:
Join from PC, Mac, Linux, iOS or Android:  https://lyrasis.zoom.us/my/dspace  (Meeting ID: 502 527 3040)

Attendees
Actual attendee list will be updated after meeting.

- Tim Donohue - LYRASIS
- Pascal-Nicolas Becker - The Library Code (unavailable)
- Alexander Sulfrian - Freie Universität Berlin
- Lieven Droogmans - Atmire
- Ben Bosman - Atmire
- Paulo Lopes - FCT|FCCN
- Mark H. Wood - IUPUI
- Paulo Graça - FCT|FCCN
- Jose Carvalho - University of Minho
- Dimitris Pierrakos - ARC/OpenAIRE
- Oliver Goldschmidt - Hamburg University of Technology (TUHH)
- Helen Baer - Colorado State University

Agenda

<table>
<thead>
<tr>
<th>#</th>
<th>Time</th>
<th>Item</th>
<th>Information</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5mins</td>
<td>Agenda</td>
<td>Any additional topics to today's agenda?</td>
<td>All</td>
</tr>
<tr>
<td>2</td>
<td>15min</td>
<td>Working to finalize task list for DSpace 7</td>
<td>Review our updated Task List (to be drafted by Lieven Droogmans)</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Estimation Process</td>
<td>• Filter out tasks that are &quot;out of scope&quot; for DSpace 7 (likely in DSpace 8 or beyond)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Define/scope/subdivide tasks that we want to estimate for DSpace 7.0 inclusion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Looking for additional volunteers to help with estimation of tasks!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Tim to report back on whether he can be the secondary estimator on all tasks.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5 mins</td>
<td>PR 74</td>
<td>Any open feedback on Entities support for external authority sources <a href="https://github.com/DSpace/Rest7Contract/pull/74">https://github.com/DSpace/Rest7Contract/pull/74</a>?</td>
<td>Ben</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>This is a dependency for the new PRs 82 and 83 and the development in REST and Angular</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15 mins</td>
<td>AIP</td>
<td>Proposal on AIP export/import <a href="https://docs.google.com/document/d/1X0Xspp2Y0tPbmq7yXwmLu7BbMAILxxOCNbw0">https://docs.google.com/document/d/1X0Xspp2Y0tPbmq7yXwmLu7BbMAILxxOCNbw0</a> ri7Y/edit?view=header&amp;hl=en&amp;gl=us</td>
<td>Ben</td>
</tr>
<tr>
<td>5</td>
<td>10mins</td>
<td>Wrap-up and Assigning tasks</td>
<td>Assign tasks / PRs to review (as needed)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Next meeting is Tues, Oct 22</td>
<td></td>
</tr>
</tbody>
</table>

Current Work

Tickets to Resolve
PRs Needing Review

1. (REST Contract) Entities support for external authority sources https://github.com/DSpace/Rest7Contract/pull/74 UPDATED (Paulo Graça, Alexander Sullivan, Tim Donohue)
2. (NEW) (REST Contract) Entities: Creating an archived item from an external source https://github.com/DSpace/Rest7Contract/pull/82 (Paulo Graça, Tim Donohue)
3. (NEW) (REST Contract) Metadata suggestions in the live import https://github.com/DSpace/Rest7Contract/pull/83 (Paulo Graça, Tim Donohue)
4. (REST) Discovery indexing: Ensuring discovery configuration is used during indexing https://github.com/DSpace/DSpace/pull/2527 (Tim Donohue, Paulo Graça)
5. (REST) Ds 4224 paginated methods for relationships https://github.com/DSpace/DSpace/pull/2535 (Chris Wiper - feedback provided, Alexander Sullivan)
6. (NEW) (REST) Entities bug-fixes in place management https://github.com/DSpace/DSpace/pull/2551 (Chris Wiper, Tim Donohue, Paulo Graça)
7. (NEW) (Angular) Name variants on item pages https://github.com/DSpace/DSpace-angular/pull/475 (Paulo Graça, Tim Donohue)
8. (NEW) (Angular) Name variants on item pages https://github.com/DSpace/DSpace-angular/pull/496 (Paulo Graça, Tim Donohue)

PRs Merged this week!

1. (REST Contract) Configuring whether name variants should be used https://github.com/DSpace/Rest7Contract/pull/77
2. (REST) (Entities) CSV Import fixes, improvements to entity validation: https://github.com/DSpace/DSpace/pull/2522

Task List

This list was roughly prioritized in the meeting on May 23, 2019 (just before OR2019). The prioritization below may change, but it gives a high level overview of what still needs to be done. NOTE: Keep in mind, just because an item is listed here does NOT guarantee it will be completed for DSpace

7. Some of these tasks may need to be delayed for a future release.

1. Submission integration (creating Entities & relations using the Item submission process) - Mockups already created by Paulo previously. - In implementation
   a. Not started yet: Adding name variants, importing from external sources
   b. Which metadata fields should be used for each Entity type. (DS-4223). - In implementation
2. Configuration of batch import (via CSV) for Entities - Already a CSV import available, but can only link entities in CSV to existing entities (in the system). Need to decide how to represent relations in CSV. - In implementation
   a. Needs analysis of what is outstanding.
3. Permissions on Relations (who has permissions to add/modify/delete relations) - Currently, if you have Edit permissions on the Entity, then you can edit/delete any relationships to/from that Entity. - NEEDS DISCUSSION
4. Deleting objects with Relations (How or should deletion propagate between closely related objects, e.g. delete entire Journal) - Currently, deleting a relation just decouples the two Entities. E.g. If you delete a Person entity, that Person may no longer be listed on any Publications it is linked to (may want to copy info over after deletion). - DISCUSSED on Aug 6
   a. Relates to GDPR
5. AIP Backup & Restore (of Entities)*
6. Dynamic display of Relations - determine automatically how a list of entities displayed on an Item page (list vs search). Currently hardcoded based on entity type (in item page template). Want to make it configurable/dynamic. (NEEDS DISCUSSION)
7. Importing from external sources (can estimate)
8. OAI-PMH configuration which maps Entity metadata fields to OpenAIRE v4 (may want to copy info over after deletion). - DISCUSSED on Aug 6
9. ORCID integration with Entities (Person Entities), (Related to #15)
   a. Importing from external sources (can estimate)
10. Best Practices around Entities in Collections. We've suggested in the Preview Release to structure Collections based on Entity Type (Person Collection, Projects Collection, etc). We should better document and formalize these best practices. Can we hide these Collections which only serve to store Entity Type.
11. The ability of pick the proper affiliation of a Person for a specific context. DSpace should address this use case to allow the user to describe something like in this document http://repository.uminho.pt/bitstream/1822/46268/1/1-s2.0-S1877050917302788.pdf regarding the authors and affiliations. You have different persons, each can belong to an institution at the time of that publication. The affiliation shouldn't be changed afterwards. And the user should be able to pick the proper one if an Author has more than one. - PER discussion on July 16, this is lower priority and may be more likely to be implemented in DSpace 8
12. (REQUIRED) Deeper dive discussion into how Entities will work with Authority Control, and how ORCID integration will work (as ORCID integrates with Authority Control). Some useful resources: Discussed in meetings on Aug 27 and Sept 17 - In implementation
   a. Authority Control of Metadata Values
   b. ORCID Integration#Importingexistingauthors&keepingtheindexuptodate
Any open feedback on Entities support for external authority sources https://github.com/DSpace/Rest7Contract/pull/74?

* Tim notes this PR should not be changing the /authorities endpoint. It's an Entities-specific PR, so it should just update *externalsources* endpoint for Entities.
* That said, if there are necessary refactoring/changes to /authorities endpoint those can be in a separate Contract PR, submitted to the DSpace 7 WG
* Ben notes that makes sense to him too
* Tim had questions in the PR on whether the "metadata" section coming back from each External Source would display DSpace schemas? That seems to imply we'd be "pre-mapping" the response of the external source to DSpace metadata?
* Ben answers Yes, we would need to do this pre-mapping to support the Angular UI anyways. We don't want the Angular UI to have to "understanding" each different response from each different external source, so we need to map them to an understandable format /structure (DIM metadata in JSON response). This simplifies things for the client. Also, this same mapping can make the import process easier once a user selects an external object to import into an Entity (as it's metadata is already mapped)
* Tim understands now. This PR looks generally good. Will add some notes to the approval though that if we hit performance issues or other issues with implementing this generic endpoint, then (as discussed last week) we may need to revisit this contract. But, it seems like a good place to start work.

Proposal on AIP export/import https://docs.google.com/document/d/1X0XspzYOottbmq7yXwxmu7FbMAfLxxOCOnbw0_r7Y/edi#heading=h.qi8bpl6kog7yi

* Ben walked us through this early proposal.
* One of the main questions were around adding the "place" field to DIM output (to represent the ordering of metadata fields, especially important for mixed metadata – relationships and string metadata)
  * Tim and Paulo note this DIM output is likely used elsewhere (Paulo thinks the OAI harvesting uses it too). But, that's not necessary a bad thing – the place field could be helpful there too
  * Generally, no one disapproves of this direction. Seems like a "place" field would help simplify things, but we may want to make it optional. If the metadata fields are already ordered, then place can be assumed to be the order they are in. But, when needed (especially for entity relationships), "place" can be explicitly added to the field.
* Other discussion notes were added to the Google Doc directly

Review our Updated Task List

* Lieven is on holiday, and he wasn't able to pass along his work on this.
* We quickly ran through the tasks listed as "in implementation" to determine if any are complete. None seem to be. So, we need to better define/determine the remaining subtasks for each in order to add to an estimation spreadsheet
* Lieven did have some notes on outstanding tasks from the North American User Group meeting talk on Entities. Tim thinks those may have been higher level, but will review to see if they can be applied to this task list to make estimations easier here.

Wrapped up meeting with assignment of PRs. Next meeting is Tues, Oct 22.