2019-10-08 DSpace 7 Entities WG Meeting

Date & Location

08 Oct 2019 at 15:00 UTC (11:00am EST)

Location:

Join from PC, Mac, Linux, iOS or Android: **https://lyrasis.zoom.us/j/9963190968** *(Meeting ID: 996 319 0968)*

NOTE: OUR MEETING LOCATION HAS CHANGED FOR TODAY ONLY! PLEASE USE THE ABOVE LINK

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

#	Time	Item	Information	Who
1	5mins	Agenda	Any additional topics to today's agenda?	All
2	10min	Update on Submission Integration questions	Proposal (from Tim & Ben) about how to resolve questions in https://github.com/DSpace/Rest7Contract/pull/74	Tim
3	15min	Working to finalize task list for DSpace 7 Estimation Process	- Review our Task List below and begin to - Filter out tasks that are "out of scope" for DSpace 7 (likely in DSpace 8 or beyond) - Define/scope/subdivide tasks that we want to estimate for DSpace 7.0 inclusion - Looking for additional volunteers to help with estimation of tasks!	All
4	10mins	Wrap-up and Assigning tasks	- Assign tasks / PRs to review (as needed) - Next meeting is **Tues, Oct 15**	

Current Work

Tickets to Resolve

- All JIRA tickets tagged with "Configurable Entities": https://jira.duraspace.org/issues/?jql=project%20%3D%20DS%20AND%20component%20%3D%20%22Configurable%20Entities%22
    - Critical Tickets: https://jira.duraspace.org/issues?jql=project%20%3D%20DS%20AND%20priority%20%3D%20Critical%20AND%20component%20%3D%20%22Configurable%20Entities%22

- Major Tickets: https://jira.duraspace.org/issues?jql=project%20%3D%20DS%20AND%20priority%20%3D%20Major%20AND%20component%20%3D%20%22Configurable%20Entities%22
- **(HIGH PRIORITY)** Bug with performance of Angular (especially with regards to Entities, e.g. in PR#402): https://github.com/DSpace/dspace-angular/issues/458

PRs Needing Review

1. (REST Contract) Entities support for external authority sources https://github.com/DSpace/Rest7Contract/pull/74 UPDATED (Paulo Graça - feedback provided ,Alexander Sulfrian, Tim Donohue)
2. (REST Contract) Configuring whether name variants should be used https://github.com/DSpace/Rest7Contract/pull/77 READY (✅ Paulo Graça, ✅Alexander Sulfrian, Tim Donohue)
3. (REST) Discovery indexing: Ensuring discovery configuration is used during indexing https://github.com/DSpace/DSpace/pull/2527 (Tim Donohue-re-review, Paulo Graça)
4. (REST) (Entities) CSV Import fixes, improvements to entity validation: https://github.com/DSpace/DSpace/pull/2522 (Ben Bosman - feedback provided, Paulo Graça - feedback provided)
5. **(NEW)** (REST) Ds 4224 paginated methods for relationships https://github.com/DSpace/DSpace/pull/2535 (Chris Wilper - feedback provided, Alexander Sulfrian - feedback provided)
6. (Angular) Item page entities changes/refactoring https://github.com/DSpace/dspace-angular/pull/475 (Paulo Graça, Tim Donohue)

PRs Merged this week!

1. ✅ (REST) Rename properties and support for name variants https://github.com/DSpace/DSpace/pull/2488
2. ✅ (Angular) Rename relationship type properties https://github.com/DSpace/dspace-angular/pull/461

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*
2. Which metadata fields should be used for each Entity type. (DS-4223). - *In implementation*
3. Additional data for relations (essentially "metadata" or labels on relations) - Related to many other features / use cases. - *In implementation*
4. Author name variants - *In implementation*
   a. *Can estimate*
5. 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*
6. 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**)
7. 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
8. AIP Backup & Restore (of Entities)*
9. 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**)
10. SWORD integration (submission of Entities via SWORD) - Uses same format as AIP. Once AIP is implemented, SWORD should be easy.
11. OpenAIRE v4 implementation using Entities* - Brought up in Steering. Possibly just an OAI-PMH configuration which maps Entity metadata fields to OpenAIRE v4
12. ORCID integration with Entities (for Person Entities). (**Related to #15**)
    a. **Importing from external sources (can estimate)**
13. 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.
14. 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://repositorium.sdum.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)*
15. **(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*

   - Authority Control of Metadata Values
   - ORCID Integration#Importingexistingauthors&keepingtheindexuptodate

Notes

- Quick intros (welcome to Helen)
- Discussion of https://github.com/DSpace/Rest7Contract/pull/74
  - Ben has removed any creation of Authority control values. This PR will be limited to Entities only
    - We also discussed that any changes to Authority Control values should be brought to / proposed to the DSpace 7 Working Group (2016-2023), as that is outside the scope of the Entities WG. We should remain concentrated on Entities alone...though if we have broader ideas/proposals we can present them to the DSpace 7 WG
  - Discussion of whether the POST to create Entities should be on the /items endpoint or /workspaceitems

- If on /items, this would be Admin only. Only Admins would be able to create Entities from ORCID or similar external systems. This endpoint also skips workflow approval (which is why it's Admin only)
- If on /workspaceitems, then more people could create Entities. However, Entities pulled from ORCID would also need to go through workflow approval (if enabled on the Collection)
- After much discussion, we've realized this is very similar to DSpace 6.x Live import feature: 2016 Framework for live import from external sources
  - This is not yet built into 7.x, but is still in our Planning spreadsheet.
- Live import seems like the same use case, as its purpose is to lookup information from an external source and then create an Item based on that external source metadata
- We agree that there seems like a dependency here on that Live Import feature first being better defined
  - Later discussion that we need to keep in mind that this contract is currently assuming that we can *generalize* different external sources under a single endpoint. While we hope this to be true, Andrea Bollini (4Science) 's comment is correct. If we begin to find that during implementation, generalization is extremely difficult or impossible, we may need to revisit this REST contract and consider separate endpoints per external service (ORCID, PubMed, etc).
  - Summary: For the time being, we will move forward with this contract as a **read-only** endpoint. Creation of Entities from external sources may depend on Live Import (see above). Any suggestions to changes to Authority system should go through the borader DSpace 7 WG. If we find during implementation that we cannot have a generalized endpoint, we will remove it and replace with specific endpoints per external source.
- Discussion of our list of Tasks
  - Lieven Droogmans has a summarized list of remaining tasks that he presented at the North American User Group
    - He has volunteered to rework the above "Task List" based on the list of remaining tasks he gathered.
    - We will then revisit this list next week to ensure it is complete & has enough detail
  - Lieven notes that getting two estimators per task might be more difficult in this smaller group. Could Tim do the secondary estimates for each task?
    - Tim says it might be possible, but would want to check with Heather too. It'd put more on her shoulders to ensure I'm also unable to see others estimates, etc. Remember, the two estimators should not be able to see each others estimates as it could introduce accidental bias to the estimates.
    - Tim will report back on discussion internally.
- Next meeting is next week, Oct 15. Main topic right now is narrowing down our list of tasks to those which will be estimated & those which we feel are likely out-of-scope for DSpace 7.