

2019-12-03 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

03 Dec 2019 at 16:00 UTC (11:00am EDT)

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](#) - LYRISIS
- [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	25mins	Updates on OpenAI REv4	<p>Any updates to discuss this week?</p> <p>Question from last week (Nov 26): How do we handle "relatedIdentifiers" vs "alternatIdentifiers" in our DSpace metadata?</p> <ol style="list-style-type: none">1. The new Schema.org identifier fields in PR#2576 seem to correspond to "alternatIdentifiers" (e.g. person.identifier.gsid is an alternative Google Scholar identifier for the Person Entity).<ol style="list-style-type: none">a. These fields represent (externally created) Identifiers that correspond to the Entity in DSpace. For example, if a publisher has its own DOI, then the publisher's DOI would be stored in these fields...while the DSpace-generated DOI would be stored in "dc.identifier.uri" as normal.2. "relatedIdentifiers" concept could be represented as new "oaire.relatedIdentifier."* fields perhaps? (Jose Carvalho and Paulo Graça will discuss and get back to us) <p>Discussion about entities on XOAI layer (https://github.com/DSpace/DSpace/pull/2592) The way XOAI uses Solr, we foresee that Solr will get some weight and also the "/dspace/bin/dspace oai import" method will be a little bit slower using this approach (this was negligible with our experience with DS5, but we just processed Publications, now we may also process Persons). At the moment the only possible solution that occurs is to have a configuration (perhaps on oai.cfg) to enable/disable entities. Other suggestions includes filtering by the relationship label, like: isAuthorOfPublication.</p> <p>Work based on:</p> <ul style="list-style-type: none">• Initial presentation: https://docs.google.com/presentation/d/1G7JGRHq-qgQ44GSr3guBEUEQTb2-6JjKsQ0-cPGkBM/edit?usp=sharing• Review of document: https://docs.google.com/document/d/1mUrtxbVMhCmp18doZ2kavZAAJ2LMemmeiYPPC6O13LU/edit?usp=sharing• Work plan: https://docs.google.com/spreadsheets/d/1SAsQYgRAyAgdSrWJAjQmJwA2LT-T0yqXrhFu4J6xhMY	All
3	15mins	Wrap-up and Assigning tasks	<ul style="list-style-type: none">• Assign tasks / PRs to review (as needed)• Next meeting is Tues, Dec 10	

Current Work

Legend for status icons

★ = Highest Priority tasks (please prioritize these reviews/tasks over others). These are tasks with lots of dependencies

✘ = review done, changes were requested or bugs found.

✔ = review done, approved.

⚠ = review done, merge conflict or other minor changes requests

1 APPROVAL = pull request only requires a single approval to merge. This is generally reserved for PRs which are either smaller, obvious, and/or bug fixes with tests to prove they 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>
- **(REST Contract)** There could be more than one Entity reference for external sources <https://github.com/DSpace/Rest7Contract/issues/90>

PRs Needing Review

1. (REST Contract) Metadata suggestions in the live import <https://github.com/DSpace/Rest7Contract/pull/83> (Paulo Graça ✔ - I'm ok with the current status of this PR, Tim Donohue - provided feedback)
2. (REST) Delete item with relationships: configure defaults <https://github.com/DSpace/DSpace/pull/2599> (Paulo Graça - feedback added, Alexander Sulfrian, Tim Donohue)
3. (REST) Creating an archived item from an external source <https://github.com/DSpace/DSpace/pull/2590> (Paulo Graça ✔ - It's Ok, but should we rely solely on the relationship.type from collection item template?, Tim Donohue)
4. **(NEW)** (REST bugfix) bug update relationships <https://github.com/DSpace/DSpace/pull/2610> (Paulo Graça ✔, Tim Donohue ✔)
5. (Angular) Keep virtual metadata on relationship delete <https://github.com/DSpace/dspace-angular/pull/530> (Paulo Graça, Tim Donohue)
6. (Angular) Create relationships during the submission <https://github.com/DSpace/dspace-angular/pull/531> (Paulo Graça - it's Ok, but still is WIP, Tim Donohue)
7. **(NEW)** (Angular) Tabbed display of relations: <https://github.com/DSpace/dspace-angular/pull/517> (Paulo Graça ✔ - shouldn't it be also available at main search?, Tim Donohue ✔) **(MERGE CONFLICTS)**
8. **(NEW)** (Angular) Item page entities improvements <https://github.com/DSpace/dspace-angular/pull/532> (Paulo Graça, Tim Donohue)
 - a. sample screenshots: [pr-532.pdf](#)
9. **(NEW)** (Angular) Virtual metadata on item delete <https://github.com/DSpace/dspace-angular/pull/533> (Paulo Graça, Tim Donohue)
 - a. sample screenshots: [pr-533.pdf](#)
10. **(NEW)** (Angular) Reordering related entities in the submission <https://github.com/DSpace/dspace-angular/pull/534> (Paulo Graça, Tim Donohue)
11. (OpenAIRE4) OpenAIRE 4 required metadata fields <https://github.com/DSpace/DSpace/pull/2576> (✔Dimitris Pierrakos, Ben Bosman - provided feedback, ✔Tim Donohue)
12. **(NEW)** (OpenAIRE4) OpenAIRE 4 submission forms and virtual metadata [WIP] <https://github.com/DSpace/DSpace/pull/2608> (Dimitris Pierrakos, Ben Bosman, Tim Donohue)
13. **(NEW)** (OpenAIRE4) OpenAIRE4 oai_openaire metadata format support [WIP] <https://github.com/DSpace/DSpace/pull/2592> (Dimitris Pierrakos, Tim Donohue, Ben Bosman)

PRs Merged this week!

1. ✔ (REST) Feature: external sources <https://github.com/DSpace/DSpace/pull/2560>
2. ✔ (REST) Entities: Projects discovery configuration <https://github.com/DSpace/DSpace/pull/2598>
3. ✔ (REST) Delete item with relationships <https://github.com/DSpace/DSpace/pull/2577>
4. ✔ (OpenAIRE4) OpenAIRE 4 Entities and Relationships <https://github.com/DSpace/DSpace/pull/2575>

Task List

This task list has been updated as of our meeting on [Oct 22, 2019](#). The tasks are numbered for easy reference, but are not necessarily listed in priority order. During this meeting we worked to re-summarize current work status so that we can align this task list with the [DSpace 7 Estimation Process](#) (and as such, estimate any we feel should be considered for 7.0 release). *NOTE: Keep in mind, just because a task 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.*

Legend for status icons

✔ = task considered "completed" (unless bugs or issues are later found)

⚠ = task is incomplete or has further work to be done.

❓ = task that may be delayed or postponed for after 7.0.

- ✔ Metadata schemas for Entity types (DS-4223).
 - REST: <https://github.com/DSpace/DSpace/pull/2443> , Angular: <https://github.com/DSpace/dspace-angular/pull/420>
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.wjaqg235p53r
- ⚠ (Minor refactor) Decide which metadata field should be used to store Entity Type (DS-4184). Currently, it is stored in "relationship.type". **(DEE MED LOWER PRIORITY, but "nice to have" for 7.0 if possible?)**
- ✔ Creating Relations (between Entities) with CSV Batch Import
 - REST: <https://github.com/DSpace/DSpace/pull/2269> and <https://github.com/DSpace/DSpace/pull/2471> and <https://github.com/DSpace/DSpace/pull/2522>
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.df1z7jh9mcc
- ✔ Permissions on Relations (between Entities)
 - Discussed on [Aug 6](#)
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.y6qarjnylex
- Deleting Entities that have existing relations
 - ✔ Deletion action itself:
 - REST Contract: <https://github.com/DSpace/Rest7Contract/blob/master/relationships.md#deleting-a-relationship> , REST: <https://github.com/DSpace/DSpace/pull/2332> , Angular: <https://github.com/DSpace/dspace-angular/pull/402>
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.c62y8iqnvlur
 - ⚠ Copy virtual metadata from deleted entity to related entity
 - ✔ REST Contract: <https://github.com/DSpace/Rest7Contract/pull/78>
 - ✔ REST Impl: <https://github.com/DSpace/DSpace/pull/2577>
 - ⚠ Angular Impl (in progress / under review): <https://github.com/DSpace/dspace-angular/pull/530>
- ✔ Dynamic display of Relations
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.pm34t6u1djdj
 - ⚠ Completed, but may require usability improvements (*NEEDS DISCUSSION - may be delayed for v7.1*)
- Submission Integration Tasks
 - ✔ Creating new Entities using Submission Forms
 - REST: <https://github.com/DSpace/DSpace/pull/2443>
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.9aixusqzgcnp
 - ⚠ Creating relations between two Entities during Submission process
 - ✔ REST Contract: <https://github.com/DSpace/Rest7Contract/pull/64> ,
 - ✔ REST Impl: <https://github.com/DSpace/DSpace/pull/2472>
 - ⚠ Angular Impl (in progress / under review): <https://github.com/DSpace/dspace-angular/pull/531>
 - Early Docs / Notes: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_rl7jY/edit#heading=h.9aixusqzgcnp
 - ⚠ Search External Sources (includes ORCID integration with Entities)
 - ✔ REST Contract: <https://github.com/DSpace/Rest7Contract/pull/74>
 - ✔ REST Impl: <https://github.com/DSpace/DSpace/pull/2560>
 - iii. **Angular is in implementation (no PR yet)**
 - ⚠ Convert External Sources to an Entity
 - If an Admin:
 - ✔ REST Contract: <https://github.com/DSpace/Rest7Contract/pull/82>
 - ⚠ REST Impl (under review): <https://github.com/DSpace/DSpace/pull/2590>
 - If a Submitter (**Feature is NOT specific to Entities, as this is simply porting the "Live Import" framework of DSpace 6.x**)
 - ⚠ REST Contract under review: <https://github.com/DSpace/Rest7Contract/pull/83>
- Name Variants
 - ⚠ Create name variants in submission
 - ✔ REST Impl: <https://github.com/DSpace/DSpace/pull/2561>
 - ⚠ Angular Impl (in progress / under review) - this is included as part of <https://github.com/DSpace/dspace-angular/pull/531>
 - ✔ Display name variants on Item pages (Angular)
- ⚠ OpenAIRE v4 support (using Entities)
 - ✔ Configuration of Entities needed for OpenAIRE v4: <https://github.com/DSpace/DSpace/pull/2575>
 - ⚠ Configuration of Metadata fields/schemas needed for OpenAIRE v4 (under review): <https://github.com/DSpace/DSpace/pull/2576>
 - ⚠ Configuration of Submission Input Forms for OpenAIRE v4 (**in implementation, no PR yet**)
 - ⚠ Configuration of OAI-PMH for OpenAIRE v4 (**in implementation, no PR yet**)
- ⚠ How to display related items on an Item page if that related item is still in workflow approval
 - Early idea is to simply display the related item as plain text metadata (until workflow approval completes).
- ⚠ Edit Item page integration
 - How to display relationships on "edit metadata" tab in UI. How to add relationships on "relationships" tab in UI.
 - Angular UI Mockups discussed/approved on [Nov 5](#) and [Nov 12](#).
- ❓ Relations in AIP Backup & Restore
 - Discussed on Oct 15**

- b. Proposal: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_r17jY/edit#heading=h.qi8bp6kog7yi
- 13. ? Creating Relations in SAF Import (might be postponed for a future release)
 - a. Requires AIP Backup & Restore
 - b. Proposal: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_r17jY/edit#heading=h.n8ktliibe7kj
- 14. ? SWORD integration (might be postponed for a future release)
 - a. Requires AIP Backup & Restore, as SWORD uses the same crosswalks / format as AIP.
 - b. Proposal: https://docs.google.com/document/d/1X0XsppZYOtPtbmq7yXwmu7FbMAfLxxOCONbw0_r17jY/edit#heading=h.xufwyy1ep8h2
- 15. ? Best Practices around Entities in Collections (NOT: this may just be early documentation). 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.
 - a. Should we eventually consider hiding these Collections which only serve to store Entity Types?

Notes

- **Question from last week (Nov 26):** How do we handle "relatedIdentifiers" vs "alternateIdentifiers" in our DSpace metadata?
 - The new [Schema.org](https://schema.org/) identifier fields in [PR#2576](https://github.com/DSpace/DSpace/pull/2576) seem to correspond to "alternateIdentifiers" (e.g. person.identifier.gsid is an alternative Google Scholar identifier for the Person Entity).
 - These fields represent (externally created) Identifiers that correspond to the Entity in DSpace. For example, if a publisher has its own DOI, then the publisher's DOI would be stored in these fields....while the DSpace-generated DOI would be stored in "dc.identifier.uri" as normal.
 - "relatedIdentifiers" concept could be represented as new "oaire.relatedIdentifier.*" fields perhaps? ([Jose Carvalho](#) and [Paulo Graça](#) will discuss and get back to us)
 - DECISION: was to remove the identifier fields from publication and rely on `dc.identifier.*` for that entity type (for now). Some identifiers were added to other Entities however. See [PR#2576](https://github.com/DSpace/DSpace/pull/2576) for updates
- **Discussion about entities on XOAI layer (<https://github.com/DSpace/DSpace/pull/2592>)** The way XOAI uses Solr, we foresee that Solr will get some weight and also the "/dspace/bin/dspace oai import" method will be a little bit slower using this approach (this was negligible with our experience with DS5, but we just processed Publications, now we may also process Persons). At the moment the only possible solution that occurs is to have a configuration (perhaps on oai.cfg) to enable/disable entities. Other suggestions includes filtering by the relationship label, like: isAuthorOfPublication.
 - Some possible OAI-PMH performance issues noted, if we are loading large numbers of objects (e.g. all related Entities) in order to fill out the response of one Entity.
 - Possible solution: Do **not** load up all related Entities...instead, just load the info available on the relationship object (i.e. a Label and the Entity ID). This may speed up OAI-PMH requests as we'll be loading less objects. However, it may not be exactly what the OAI harvesters expect (uncertain).
 - Could also consider precaching these responses for speed, but that may add a lot of complexity to DSpace itself. The primary limitation here is within OAI-PMH, and this isn't a DSpace-specific issue.