

Metadata WG Team meetings

New doc for 2024: [Metadata WG Meetings 2024](#)

December 14, 2023

Agenda

- **Key takeaways from “rotation preferences” survey**
 - Develop clear tasks assigned to individuals and a roadmap with deadlines.
 - Provide alternative methods of engagement within meetings (e.g. voting/Menti poll, Miro boards).
 - Have time-limited, smaller group discussions for specific tasks.
 - With well-defined goals for 2024, we can introduce this without having the overhead and division of formal “sub-groups”.
 - Clarify responsibilities for group participation, including commenting on proposals and taking on assigned action items.
 - Consider involving the Metadata WG in metadata analysis tasks, such as use of existing metadata beyond required fields.
- **Tasks/goals for 2024**
 - [Metadata WG 2024 Planning](#) (work in progress)
 - We will try using GitHub issues to break down tasks and assign them to individuals/small groups.
- **Finalizing revised prioritization and inclusion criteria for controlled vocabularies**
 - [Controlled Vocabularies: Prioritization and Inclusion Criteria](#)
 - [Controlled Vocabularies: Workflow and Next Steps](#)
- **Identifying comparable vocabularies**
 - [DataCite Controlled Vocabularies: Comparators](#)
- **Update on Schema 4.5**
 - Anticipated release date - mid-January 2024
 - “Publisher” attributes required structural change to how the Metadata Schema is expressed in JSON (“DataCite JSON”) - from a string → object
 - REST API will not have breaking changes, workaround:
 - The REST API will accept both string and object values
 - To expose the publisher attributes, “publisher” parameter (=true) added similar to “affiliation” parameter that was added for Schema 4.3
- **FYI: Schema 3 deprecation timeline**
 - Announcement will go out on December 20th, 2023.
 - Email campaign to all members in January.
 - Schema 3 will be deprecated on January 1, 2025.

08:00 UTC Notes

Attendees: Isabel, Jessica, Anu, Vanessa

-We all agree that it will be useful to set clear tasks and priority topics for discussion with a clear, and not too long time framework, to avoid losing track of where we stand in terms of pending issues. Further, more information about the proposed activity of WG members around metadata análisis tasks is welcome

-Planning for 2024 gdoc looks good in general; we discussed the convenience to state clearly in the upcoming RFC document to what version of the schema the proposed changes would belong, so as to avoid confusion/false expectations in the broader community. However, it is good to request feedback from the community about “difficult to decide” topics well in advance

-We think the document around prioritization and inclusion criteria for controlled vocabularies is not yet mature enough to be shared openly, as there are many aspects to be agreed upon about the workflow and clarification of some criteria is needed (eg some seem more objective than others, and there might be some overlapping across a few of them). We also discussed about the pros and cons of sharing the whole criteria document broadly as some aspects may refer to internal management mostly

-We went through the [comparators Excel document](#) and started to fill in, it is a good and useful resource

-As regards deprecation of version 3 we wonder how many users are still in that version

16:00 UTC Notes

Attendees: Jan, Kelly, Anne, Wendy, Alex, Mike, Li, Joe

- Anne: “Tiger teams” - small groups that move quickly and present something that we can use as a strawman.
- Joe: For controlled vocabulary terms that are derived from other vocabularies, we need stability - makes sense to have our own PURL.
- Anne: If an authority is providing a controlled vocabulary, you don’t want to duplicate it or redefine things locally.
- Joe: We need to make sure our system keeps working if there are changes to the external vocabulary, but that doesn’t mean we have to redefine the term.
- Wendy: Might depend on what the term is. For example, resourceTypeGeneral is very baked into DataCite’s structure. Others like contributorType may not be.
- Joe: The other issue is whether DataCite wants to use part of a controlled list, rather than the complete list.
- Joe: Useful software: <https://www.cordra.org/>

November 9, 2023

Agenda

- Welcome Sara!
- Controlled lists editorial process: **Controlled Vocabularies: Editorial Process**
 - Proposed criteria
 - Next steps
- Controlled lists additions:
 - [relatedIdentifierType: RRID](#)
 - [relationType: IsTranslationOf/HasTranslation](#)
 - [contributorType: Translator](#)

16:00 UTC Notes

Attendees: Kelly, Wendy, Kirsten, Jessica, Sara, Joe, Mike, Alex, Li, Ted

- Wendy and Kelly will be at the DLF forum next week: <https://forum2023.diglib.org/>
 - Joe shared this resource on PIDs in heritage collections: <https://tanc-ahrc.github.io/PIDResources/>
- Controlled lists editorial process: **Controlled Vocabularies: Editorial Process**
 - Do we know of similar documents for other vocabularies/groups?
 - Joe: Getty editorial guidelines: <https://www.getty.edu/research/tools/vocabularies/guidelines/index.html#at>
 - Jessica: similar in Finnish: https://www.suomidigi.fi/sites/default/files/2020-06/JHS175_liite1.pdf
 - Will we publish this document and share it with a license (i.e. CC-BY)?
 - Joe: would be interested in reusing this
- Discussion of criteria
 - Relevance to the DataCite Community
 - *Anticipated adoption*: We discussed that this may not be as simple as the number of organizations or members—for example, a single organization registering many DOIs could represent significant anticipated adoption. The idea that the term should not have “too limited” use, which can’t be boiled down to the number of users.
 - *Domain specificity*: This is part of anticipated adoption, which has several facets, and can be combined with this factor. The group also discussed concern about “opening the floodgates”: if we allow one very specific/niche term, then this creates a precedent for similar terms to be included and could result in large, overly complex vocabularies.
 - Suitability for the controlled vocabulary

- It was raised that the “hierarchical relationships” is part of “uniqueness within the list”.
 - Test of criteria with IsTranslationOf/HasTranslation
 - We went through some of the criteria to see how they would apply and how we might use them to evaluate new terms.
 - [Controlled Vocabularies: Editorial Process - Testing Template](#)

08:00 UTC Notes

Attendees: Jan, Sara, Vanessa, Anu

Introductions: We welcomed Sara, the new DataCite metadata specialist to the meeting!

Controlled lists editorial process: Discussion around creating our own DataCite PURLs for terms available in other vocabularies. Using our own terms/PURLs is easier to govern and maintain. SKOS can provide e.g. ‘sameAs’ links.

Controlled list additions:

RRID: Agree that we need to clarify the format of the RRID - no ‘/’ in any examples but the format seems to be RRID followed by a colon and the number e.g. RRID:AB_12345

IsTranslationOf/HasTranslation: Agreed with the proposal.

contributorType:Translator: Agreed on the definition by ChatGP/Jo Padfiled because it includes software.

Actions: All to review the controlled list editorial process doc and provide comments.

All to fill in the metadata WG rota and add feedback on ways of improving communication and work of the Group.

October 30/31 2023

Agenda

1. Link to Miro board:
 - https://miro.com/app/board/uXjVNWwaoos=?share_link_id=326459589752

October 30 14:00 UTC notes

Attendees: Mike, Anne, Isabel, Alex, Li, Wendy, Kirsten, Kelly

- See notes on Miro board:
 - https://miro.com/app/board/uXjVNWwaoos=?share_link_id=326459589752


October 31 0:00 UTC notes

Attendees: Matthias, Wendy, Ted

- See notes on Miro board:
 - https://miro.com/app/board/uXjVNWwaoos=?share_link_id=326459589752

October 19, 2023

Agenda

1. Controlled lists process
 - a. Doodle poll for group to define editorial process:
<https://doodle.com/meeting/participate/id/b89oy9Ld>
 - b. Technical implementation in process with PURLs
 - c.  Schema process proposal: Controlled lists
 - i. Revised to reflect current state / progress with pipeline
2. Controlled lists priority changes
 - a. Preliminary definitions to start discussion (very drafty!)
 - b. relatedIdentifierType (GitHub discussion:
<https://github.com/datacite/metadata-wg/discussions/32>)
 - i. **RRID**: Full name: Research Resource Identifier
 1. The Resource Identification Portal was created in support of the [Resource Identification Initiative](#), which aims to promote research resource identification, discovery, and reuse. The portal offers a central location for obtaining and exploring **Research Resource Identifiers (RRIDs)** - persistent and unique identifiers for referencing a research resource. A critical goal of the RII is the widespread adoption of RRIDs to cite resources in the biomedical literature and other places that reference their *generation or use*. RRIDs use established community identifiers where they exist, and are cross-referenced in our system where more than one identifier exists for a single resource. Some examples are shown below, which are linked to metadata about each resource.
 2. Description: The Research Resource Identification Initiative provides RRIDs to 4 main classes of resources: Antibodies, Cell Lines, Model Organisms, and Databases / Software tools.
 3. Example: `<relatedIdentifier relatedIdentifierType="RRID" relationType="IsCitedBy">RRID:AB_262044</relatedIdentifier>`
 - c. relationType (GitHub discussion:
<https://github.com/datacite/metadata-wg/discussions/33>)
 - i. **IsTranslationOf**: Indicates A is a translation of B
 - ii. **HasTranslation**: Indicates A has a translation B
 - d. contributorType (GitHub discussion:
<https://github.com/datacite/metadata-wg/discussions/34>)

- i. **Translator:** A person, organization, or machine who translated the content of the resource.
 - ii. MARC relator - <https://id.loc.gov/vocabulary/relators/trl.html> A person or organization who renders a text from one language into another, or from an older form of a language into the modern form.
 - iii. From ChatGPT: Translator: An individual, entity, or automated system responsible for converting the content of a resource from one language or mode of communication into another, preserving its meaning and intended message.
- 3. Creator/Contributor roles
 - a. Miro open: https://miro.com/app/board/uXjVMjB1URs=?share_link_id=413335133164
 - b. Share any use cases for CRediT in repositories: <https://github.com/datacite/metadata-wg/discussions/7>
- 4. AOB
 - a. Call for WG members to be distributed shortly
 - b. Please fill out survey with rotation preferences by November 1: <https://forms.gle/brMs7AUdkSSLfWhM8>
 - c. Complete strategic planning survey: <https://mhcgai7tmi0.typeform.com/to/G1ADfnFe> - by November 7, 2023

8:00 UTC Notes

Attendees: Jan, Jesssica

Apologies: Kirsten

1. **Controlled list process:** Jessica and Jan are interested in joining the group to define the editorial process. Both agreed that PURLS are authoritative and well maintained. Their use by e.g. DCMI is indicative of their future longevity.
2. **Controlled list value additions:**
 RRID: Agreed that this is a useful addition.
 relationType: isTranslationOf/hasTranslation: Agreed that this is an essential addition
 contributorType= Translator: Would seem to work well beside the new relationType
3. **Creator/Contributor roles**
 We reviewed the Miro board focusing on the comparison of the DataCite contributorType values and the CRediT values.
 Repositories do use CRediT values and there are attempts to map them, though there are problems.
 A ground up revisit of contributor types would be beneficial since some of the DataCite values seem to be administrative (Sponsor; RightsHolder) or attempting to offer a pragmatic solution to a problem e.g. recording and noting that an organisation is the hosting repository (HostingInstitution).
 Given the issues with the current DataCite values it may be useful to develop a new conceptual model e.g. relatedPerson with associations or attributes. This model should

be able to map to other vocabularies. See the CSC metadata model for e.g. relatedEntity - Person/contributor role etc. <https://tietomallit.suomi.fi/model/mrd/RelatedEntity/>
One of the reasons for the WG looking at contributor types was the fact that there are no 'types' for creator e.g. editor. We also noted that creators such as software developers are not often included as creators (CRediT includes a value for 'software').

End of Meeting

15:00 UTC Notes

Attendees: Isabel, Kelly, Anne, Wendy, Ted, Alex, Mike, Joe, Jan

1. Controlled lists update
 - a. Poll is for first meeting
 - b. Please complete by **Friday, Oct 20**
 - c. Deliverable: To agree on the process for updating controlled vocabularies. *Not* an editorial group itself. Working on how we manage vocabularies.
 - d. Joe: decision made for vocabulary server for E-RIHS - Opentheso: <https://opentheso.hypotheses.org/>. However, it requires a server (would not work for DataCite's approach at this time). GitHub repository: <https://github.com/miledrousset/Opentheso2>
2. Controlled lists priority changes
 - a. relatedIdentifierType: RRID
 - i. Current relatedIdentifierType list: <https://datacite-metadata-schema.readthedocs.io/en/4.5/appendices/appendix-1/relatedIdentifierType/>
 - ii. Ted: 650,000 RRIDs have been registered. Discussion with Anita (lead for RRID at UCSD) has been about using RRID for repositories.
 - iii. Anne: How do we decide which identifiers to recognize in the relatedIdentifierType list? There are many; how do we manage this complexity?
 - iv. Joe: What are the joining criteria?
 1. Is it the quantity of an identifier being registered? How is it being used?
 2. Wendy: stability of the identifier scheme?
 - v. Joe: Note this is focused on biology and life sciences - antibodies, cells, etc.
 - vi. Ted: group at UCSD is working to extend the breadth of RRIDs. These are being referenced from papers. Would lean towards being permissive in the DataCite schema.
 - vii. Anne: think about system impact. What support do we provide for local identification systems? Identifiers overlap (e.g. organization IDs,

- repository IDs). what happens when ID organizations disappear? How much does DataCite care?
- viii. Kelly: what criteria were looked at previously by the WG?
 1. Isabel: weren't so many IDs years ago - was easier to be more inclusive.
 2. Anne: requests have been handled as part of a schema update - requested by community, reviewed by Metadata WG, included in schema release.
 - ix. Ted: benefits being able to connect to various kinds of resources from DataCite metadata - benefits DataCite user community. Already have domain-specific identifiers (LSID, PMID). note there is nothing called Other in relatedIdentifierType list. Trivial amount of overhead in this list.
 - x. Jan: thought adding RRID was a no-brainer this morning. This is widely used in biomedical sciences, authoritative. Emphasis on the flexibility of the schema.
 - xi. Joe: comes down to administrative burden. If we're not saying an identifier is endorsed/connected to DataCite, it is just another identifier and we are allowing people to say what type that is. If there is recognition, interoperability, etc. that is more complicated
 - xii. Kelly: longer term we do need to be thinking about accuracy and validation, possibly how we clean and curate metadata. Short term: looking at defining RRID.
- b. **relationType: IsTranslationOf / HasTranslation**
- i. Kelly: longstanding request from many users. Workarounds are IsDerivedFrom/IsSourceOf, IsVariantFormOf? - not used consistently
 - ii. Joe: machine translation increasingly prevalent - as is editing (within same language) - tools like DeepL do both
 - iii. Isabel: can an organization be a translator?
 1. Wendy: yes - company
 - iv. **IsTranslationOf**: Indicates A is a translation of B
 - v. **HasTranslation**: Indicates A has a translation B
 - vi. Discussion of chains of translations between multiple languages. Isabel: relationTypes always deal with 2.
 - vii. Joe: is there any problem with the word "translation"? Are any mappings between language structures not "translations"?
 - viii. Anne: distinction from transliteration
- c. **contributorType: Translator**
- i. Draft definition: Translator: *A person, organization, or machine who translated the content of the resource.*
 - ii. From MARC relator terms: *A person or organization who renders a text from one language into another, or from an older form of a language into the modern form.*
 - iii. Machine or software? Service?

1. Joe, from chat GPT: Translator: *An individual, entity, or **automated system** responsible for converting the content of a resource from one language or mode of communication into another, preserving its meaning and intended message.*
 - iv. Anne: the translator would be cited in the translated resource (not the source) - definition should clarify this.
 - v. Kelly: is it an issue to include software, system, service in Contributor vs RelatedIdentifier?
 - vi. Joe: think in terms of role - now the role of the translator can be carried out by a software system
 - vii. Anne: the issue is the schema structure - requires a name for Contributor, nameType personal or organizational. (Do we need a new nameType?)
 - viii. Joe: aspect of how we expect people to use this. We could just say "person or organization" and blank nameType.
 - ix. Isabel: machine translation is prevalent
 - x. Wendy: how often would a DOI be assigned to something translated by DeepL?
3. Creator/Contributor roles
 - a. Reminder that Miro is still open - share use cases

Meeting chat:

08:03:00 From Kelly Stathis To Everyone:

https://docs.google.com/document/d/1WIHn9RkgxyfwJVQJvbDCc5AQM8_EEQaJ_I2uNcDEqS0/edit#heading=h.sawwpf14twqw

08:09:14 From Kelly Stathis To Everyone:

sorry for messing up the notes placement, they're now below the agenda where it says

15:00 UTC

08:10:35 From Joe Padfield (The National Gallery - UK) To Everyone:

<https://opentheso.hypotheses.org/>

08:13:41 From Joe Padfield (The National Gallery - UK) To Everyone:

<https://github.com/miledrousset/Opentheso2>

08:18:41 From Wendy Robertson (she/her) To Everyone:

stability of the identifier scheme?

08:19:01 From JAshton To Everyone:

Reacted to "stability of the ide..." with 👍

08:20:52 From Wendy Robertson (she/her) To Everyone:

This is how it appears in JATS -

<https://jats.nlm.nih.gov/publishing/tag-library/1.3d1/element/resource-id.html>

08:27:29 From Joe Padfield (The National Gallery - UK) To Everyone:

Unrelated, but just sharing for interest - just came across a tool for Documenting contributorship with CRediT: <https://rollercoaster.shinyapps.io/tenzing/>

08:34:23 From Ted Habermann To Everyone:

<https://support.datacite.org/docs/link-checker>

08:37:00 From Ted Habermann To Everyone:
Might be an interesting blog post:

08:37:06 From Ted Habermann To Everyone:
<https://upstream.force11.org/building-a-community-of-practice/>

08:38:21 From Kelly Stathis To Everyone:
So far none - <https://api.datacite.org/does?query=contributors.name:ChatGPT>

08:43:31 From Kelly Stathis To Everyone:
Variant form might actually be a better fit for transliteration? (Not a linguist either)

08:46:52 From Wendy Robertson (she/her) To Everyone:
Translator [trl]
A person or organization who renders a text from one language into another, or from an older form of a language into the modern form.

08:46:53 From Alex Whelan To Everyone:
<https://www.loc.gov/marc/relators/relaterm.html>

08:46:57 From Wendy Robertson (she/her) To Everyone:
<https://www.loc.gov/marc/relators/relaterm.html>

08:47:02 From Joe Padfield (The National Gallery - UK) To Everyone:
Translator: An individual, entity, or automated system responsible for converting the content of a resource from one language or mode of communication into another, preserving its meaning and intended message.

08:47:44 From Alex Whelan To Everyone:
Here's the marcRelator with URIs

08:47:45 From Alex Whelan To Everyone:
<https://id.loc.gov/vocabulary/relators.html>

08:47:58 From JAshton To Everyone:
Translator [trl]
A person or organization who renders a text from one language into another, or from an older form of a language into the modern form.

08:54:30 From Wendy Robertson (she/her) To Everyone:
It looks like ChatGPT is in LC authority file as a name -
<https://id.loc.gov/authorities/names/n2023013228.html> -- Deepl is not yet there

08:55:44 From Kelly Stathis To Everyone:
Reacted to "It looks like ChatGP..." with 👍

09:02:06 From Joe Padfield (The National Gallery - UK) To Everyone:
<https://www.dariah.eu/activities/open-science/data-re-use/>

September 21, 2023

Agenda

1. Controlled lists update
 - a. Pipeline in development that converts from [Google Sheet](#) to different formats

- b. Uploading to BioPortal which serves SKOS, CSV, and RDF/XML.
 - i. Working with the CEDAR team on publishing vocabularies, following from this work: <https://bioportal.bioontology.org/ontologies/DATACITE-VOCAB>.
 - ii. Test/demo for 4.5 (currently private):
<https://bioportal.bioontology.org/ontologies/DATACITE-TEST>
 - c. Working on PURLs for each vocabulary term for 4.5
 - i. DataCite now has the purl namespace /datacite
2. Summary and discussion of draft RFC feedback
- a. [Clarify supported language codes](#)
 - i. “Allowed values, examples, and other constraints” and the accompanying footnote were updated to incorporate feedback.
 - ii. Suggestion to capture scheme URI has been noted, with challenges:
 1. We want to encourage use of the IANA registry as the first choice, with other lists used only when necessary.
 2. A given tag can have more than one source - language tags in the IANA registry overlap with those in ISO 639-1/2/3.
 3. An open-ended (non-controlled) schemeURI that is not used consistently is not machine actionable. Note that DataCite services are currently designed around ISO-639-1 two-character codes for features like language faceting, aggregations, etc. Even if services are enhanced to accommodate a new range of codes, DataCite would not be able to rely on an uncontrolled free-text schemeURI attribute to accomplish this. This would be a similar challenge for aggregators.
 4. For languages that cannot be represented by the tags in the IANA registry, an open-ended “languageInformation” sub-property may be an option (this would also be a breaking change).
 - b. Creator/Contributor roles
 - i. This proposal followed from the work of the contributor roles subgroup from 4.5. Major challenges:
 1. Removing contributorType and replacing it with an uncontrolled field, without validation, will result in less consistent metadata (even if the recommendation for that field is to use certain controlled terms).
 2. The “contributor types” are a different concept than “contributor roles” and can’t easily be combined.
 - ii. Miro board for brainstorming:
https://miro.com/app/board/uXjVMjB1URs=?share_link_id=122698801243
 - c. Coverage dateType
 - i. Based on feedback from the DataCite engineering team, looking to prioritize non-breaking change (dateType) in the near term over a solution involving deprecating GeoLocation.

3. Timelines

- a. Schema 4.5 aiming for release by year end - development work has started
- b. Next RFC sometime after schema 4.5 release (not October)

8:00 UTC Notes

Attendees: Isabel, Matthias (very late!)

Main points of discussion:

-Regarding vocabularies management, in light of ongoing DataCite work, we mentioned different tools to manage them (Bioportal, CEDAR templates, iqvoc, Vocbench, poolparty, and Research Vocabularies Australia which also embraces the management of vocabularies outside of Australia). A dedicated session about DataCite ongoing work and role of WG in this regard may be beneficial

-Governance issues related to management, maintenance and building of vocabularies: it is a good opportunity to explore collaboration with other initiatives working on vocabularies. As an example, RAID is working on its metadata schema where DataCite controlled lists are present: ongoing work, not final at <https://raid-metadata.readthedocs.io/en/latest/index.html>

-Around creator/contributor discussion: we both tend to think that the separation between creator and contributor properties may appear artificial nowadays. The relation between different contributors is the fundamental linkage around which the discussion may evolve. Matthias shared ongoing RAID work in this regard, too, with contributor subproperties as a mechanism to give due context,

<https://raid-metadata.readthedocs.io/en/latest/core/contributors.html#contributor>. Also worth mentioning RAID ongoing work around relatedObject

<https://raid-metadata.readthedocs.io/en/latest/core/relatedObjects.html#relatedobject-type-id>.

We also discussed the tricky management of research groups as a type of contributors, due to their changing nature and composition over time: in this sense, having research projects instead may be more manageable

-Also, we discussed the opportunity to think about potential use cases to justify and support changes for creator/contributor in the schema: eg, what use case is behind including CREDIT, and/or any other controlled lists in the schema?. We also commented on challenges derived of introducing breaking changes in the schema as the goal is to enable interoperability with a broader community and at the same time avoid the risk of losing existing DataCite users on the way

-PIDfest is coming in 2024! More news to come.

15:00 UTC Notes

- Vocabularies
 - Joe interested in experimenting with pipeline

- Should they be separate vocabularies with their own versioning, or one ontology?
- Need to define editorial process
- Tagging releases in GitHub too
- Is there a way just to do this in GitHub without Google Sheets?
- Multiple parent terms for Other? - currently implemented as different terms with different identifiers, same label
- Idea to introduce additional meeting - full metadata WG - in addition to monthly general calls, focused on controlled vocabulary maintenance
- Look at sustainability of bioportal
- Contributor roles
 - Could a controlled list be constructed from the children of multiple things? I.e. concatenate DataCite list and CRediT list
 - Annotation for origin of term - practically a flat list (even if there is external indication of hierarchy)
 - Need to understand use case / reason for supporting CRediT better
 - Mixing our own vocabulary and external one is messy

August 10/17, 2023

Agenda

1. Draft RFC: [DRAFT DataCite Metadata Schema Changes: Request for Comments ...](#)
 - a. Internal deadline for feedback from Metadata WG and DataCite team: September 14
 - i. Action: Please add xml example for language representation
 - b. Aiming to launch for public comment in October (will make fresh copy)
2. Controlled lists:
 - a. [Schema process proposal: Controlled lists](#)
 - b. Feedback on test repository: <https://github.com/datacite/controlled-lists/issues>
 - c. Patch versioning approach
 - d. DataCite 4.4 purls - <https://bioportal.bioontology.org/ontologies/DATACITE-VOCAB/?p=summary>
3. ResourceTypeGeneral review
4. AOB
 - a. DataCite hiring Metadata Specialist for PID4NFDI project: <https://datacite.org/jobopportunities.html#head2>

August 17th 8:00 UTC Notes

Attendees: Isabel, Matthias, Vanessa, Anu, Jan

Main issues discussed: (Other than what is already reflected in the documents and were briefly commented during the meeting)

-Language property changes:

Matthias to include new examples in the proposed documentation about language codes registries

Anu raised up the issue as regards the subproperty where to make reference to the language registry being used

-Temporal coverage changes:

Anu mentioned to make it optional and propose a standardized way to represent this new date type

-Creator/contributor changes:

Matthias shared ongoing discussion as regards RAID metadata schema:same problems as regards "roles" and important to distinguish from specific contributors to a specific resource: role would be mostly referring to the job of a person

This issue is bigger than us and requires an agreement between different communities: orcid, CROSSREF, datacite...

Anu commented that elaborating use cases for creator and contributor properties may facilitate our work advance

Jan commented on the discussion the week before

We agree that mixing vocabs about roles and types of authors may lead to confusion

-PURLS discussion

We discussed about the stability of this system to generate PIDs for the controlled values. We agreed it is a good solution

-Proposal for updating DataCite controlled lists:

It is a big enterprise for the WG, we will need a dedicated task force

to maintain the vocabularies. Anu proposed a template to encapsulate the proposals in a standard way

Isabel. what about choosing a tool that serves to publish the vocabs and also manage all the internal work by staff and WG?

All found the topic as a good opportunity for the schema but also raises concerns about maintenance, effort and control of consistency

August 10th 15:00 UTC Notes

Attendees: Anne, Kelly, Kirsten, Jan, Ted, Mike, Wendy, Joseph, Alex

- Metadata WG to comment on RFC
 - Anne: Table format has reached its limit
 - Joseph: Could we comment on the webpage?
 - <https://usersnap.com/lets-schedule-a-demo>
 - May be beneficial for commenting on the entire schema
 - Kirsten: For titles, can we be transparent about how these are indexed (or not) in DataCite? I.e. that subtitle is not combined with main title
 - Kelly to look into this - Kirsten had talked with Cody
 - Possibly add a footnote to show the importance of main title and that translated/alternative/etc aren't indexed in the same field
 - Creator / Contributor roles:
 - Discussion around merging the properties. Confusing when to use Creator vs Contributor (vs Publisher).
 - Could be difficult to transition
 - Concern that aggregators wouldn't look in Contributors for authors, publishers.
 - What about adding a creatorType - author, editor? (could we align this with roles in citeproc?)
 - Anne: what obligation does DataCite have to align their schema with other registration agencies?
 - Wendy: we should include examples for Creator/Contributor
- Controlled lists
 - Kirsten: ensure RDF has definitions. Also the "definitions" in general are not definitions.

June 8, 2023

Agenda

1. Update on 4.5

- a. Modified scope to support faster release
- b. Distribution property will be in subsequent version (4.6, 5.0)
- c. See updated documentation here:
<https://datacite-metadata-schema.readthedocs.io/en/4.5/>
2. Current proposals:
 - a. Language property expansions
 - i. [GitHub discussion](#)
 - ii. [Proposal](#)
 - iii. Note - COAR recommendations open for comment through June 30:
<https://comments.coar-repositories.org/>
 - b. Temporal coverage
 - i. [GitHub discussion](#)
 - ii. [Proposal](#)
 - iii. Discussion - which option?
3. [Publisher guidance](#)
4. Embargo date guidelines

08:00 UTC Notes

Attendees: Matthias; Kirsten; Jessica and Jan

Regrets: Anu

- Update on 4.5: some concern expressed about the Distribution/contentURL property (now moved to 4.6 release) around PIDs for the individual files.
- Language proposal: the use of xml:lang for the language of the metadata needs improved documentation. Positive that the repeatability of the language property now allows for multiple languages, not just the primary language. However, there is still concern that many aggregators will only pick up e.g. the first described language as a 'primary' language. This is especially important for bi-lingual countries or materials.
- Language proposal: it would be useful to note the flexibility of the language codes that can be used to include e.g. indigenous languages. Suggest including e.g. austlang
- Language proposal: It would be useful to have a simple way of recording which language standard or coding system was used e.g. by adding a type attribute. This would be simpler than adding e.g. LanguageScheme/SchemeURI and valueURI etc. which might potentially be discussed for inclusion in 4.6 or 5.0.
- DataCite schema version used by repositories: Do we have metrics on which schemas repositories are using (see note below)
- Temporal coverage: suggest using option 1a with dateType='coverage' and also adding dates to geoLocation.
- Publisher guidance: agree that it should be more flexible and not structured around resourceTypeGeneral.
- Embargo date guidelines: agree that these should be amended to reflect e.g. material such as theses and dissertations where the data of issue is the date the degree is awarded (or the DOI is registered) even though there is an embargo date.

Schema Versions (Ted): All DataCite API queries include a schemaVersion facet with counts for the ten most common values. A query for resourceType = 'Dataset' with 13.4 million results gave these numbers: Schema 4 (10,746,836), Schema 3 (2,169,585), Schema 4.3/ (453,781), Schema 2.2 (40,935), Schema 4.0/ (5,430), Schema 4.4 (1,928), Schema 4.3 (930), Schema V1 (471), Schema V2 (88), Schema V3 (26). The most common version (Schema 4 with no sub-version) accounts for almost 75% of all values.

For 'Text': Total: 9,736,111, Schema 4 (8,391,425 78%), Schema 3 (1,195,848), Schema 2.2 (144,632), Schema 2.1 (4206)

15:00 UTC Notes

Attendees: Isabel, Kelly, Anne, Joe, Felix, Wendy, Mike, Ted

- Anne - tips for posters
 - Research poster creation video:
https://www.youtube.com/watch?v=SYk29tnxASs&ab_channel=MikeMorrison%2CPhD
 - Template: <https://osf.io/6ua4k/>
- Update on 4.5
 - In general: Moving to smaller, more frequent versions and have complex issues for major versions.
 - Continuing to work on the current priority items - some may end up in 5.0, some may end up in an earlier version (e.g. 4.6).
 - Discussion about the support DataCite provides to earlier versions of the schema
 - Discussion of whether 5.0 will be incremental (with some breaking changes) or full redesign. Currently, not looking at a full redesign/complete overhaul but we may want to start planning for this long term.
 - Conversations about consistency - related to rebuilding the schema. In this regard, the idea of an AI based research project looking into the schema design building was mentioned
 - What about RDF?
 - Anne: Yes. In fact, the FAIR Principles want to see metadata in a format like RDF or OWL or JSON-LD - not JSON or XML. Odd what you find when you read the details...
- Language:
 - Language property:
 - Definition/guidance changes - which codes can be used in this field.
 - Expanding beyond recommending ISO-639-1 in the guidance. This is a documentation change only, there is no restriction currently.
 - Other standards should be accepted too - as long as there's a way of finding out what that tag means and there is full documentation. How can we provide a mechanism for this? What level of

complexity will the schema require to promote usage of other standards?

- Repeatability
 - 0-1 → 0-n
 - Breaking change; no earlier than 5.0.
- xml:lang attribute:
 - Can be listed explicitly as sub-property (attribute); see notes from previous discussion here:
<https://github.com/datacite/metadata-wg/discussions/16#discussioncomment-5272798>
- COAR recommendations: there are several mentions to DataCite schema v4.4 across the document and it is worth taking a look and comment about. All WG members welcome to comment on these recommendations, too
- Coverage
 - What do we mean with “coverage”? Is it clearly defined in the documentation? Different people understand different things and applications and such ambiguity makes it hard to agree what a coverage property may pack
 - geoLocation is not the same as spatial coverage: geoLocation is earth-centric, and can't refer to points in space. In DataCite schema geoLocation is not much used for spatial data
 - Temporal coverage - may not always be an exact date; may be very old - how to define degree of certainty?
 - dateInformation?
 - Ranges are supported in the Date property
 - For option 2 / coverage:
 - Suggestion to have the option to add text description of coverage, to give context. It may help reduce the existing fuzzyness around “coverage” in general and “temporal coverage” in particular
 - With Option 1a, concern about giving repositories multiple ways to express the same information.

May 11, 2023

Agenda

1. Review / discussion of scope for 5.0
 - a. See [What's next for the DataCite Metadata Schema \[Metadata WG\]](#)
 - b. Results from survey: most/all respondents interested in controlled lists subgroup
2. Brainstorm: controlled lists approach
 - a. Initial considerations:
 - i. Being able to update lists separately from schema
 - ii. Easier maintenance
 - iii. Social - community process - how to discuss changes, see proposals

- iv. Technical - DataCite systems, ways to share list across Fabrica/APIs/Commons/etc
- v. Definitions
- vi. Flat vs hierarchy
- vii. Linked data
- viii. Alignment with other/external vocabularies
 - 1. Mapping to other vocabularies (e.g. resourceTypeGeneral)
 - 2. Referencing other vocabularies (e.g. contributorType)

08:00 UTC Notes

Attendees: Jan, Anu, Kirsten, Vanessa, Matthias

Citations:

- Agree emphasis on citation as core DataCite metadata function.
- The importance of the resourceType should be made more prominent and mandatory as part of the citation e.g. for samples etc. and for interoperability with other schemas such as CrossRef.
- Could the DataCite doi API citation service include the resourceType as a mandatory element?
- Agreed that the elements required for each resource type is more important than e.g. citation style or format (which is often dependent upon e.g. external journal requirements).
-

Controlled vocabularies:

- Discussed the possibility of using relatedIdentifiers to other controlled vocabulary content.
- Discussion on decoupling the updating of the vocabularies from the publication of the main metadata schema. Pros: schemas can take a lot of time which holds back inclusion of new terms. Cons: new terms may cause validation problems for repositories against the current or older schema version.
- Maintenance of the vocabularies could be handled via the Metadata Working Group / Sub-Group rather than by a separate Vocabularies committee.
- The DataCite controlled vocabularies should be registered and made available, including as RDF linked data (not required by the main metadata schema).

RelatedItem

- Emphasis on using relatedIdentifier as default unless no doi or other identifier is available.
- Look at the work the RDA collections working group is doing on citing identifiers.

15:00 UTC Notes

Attendees: Kelly, Joe, Ted, Wendy, Alex, Kristian

Controlled vocabularies

- Joe: At National Gallery, working on setup vocabulary server.
- Joe: In between / simpler solution - using a GitHub system - tsv - GitHub actions to reformat into what's needed to ingested into your system
 - <https://github.com/E-RIHS/controlled-lists>
 - Structure includes "Same as", "source" for referencing other standards
 - Actions turn the tsv file into a blob of JSON needed for other system
 - Resources: <https://github.com/E-RIHS/hs-interoperability/tree/main/Vocabulary>
- Ted: Question is - where do you want to be between: a list of words, a list of words with definitions and possibly multiple languages, richer semantic vocabulary (e.g. SKOS)
- Joe: 2 pieces - hierarchical vocabulary vs controlled lists. Could use "parent" column in tsv file
- Ted: vocabulary server example:
https://www.bodc.ac.uk/resources/products/web_services/vocab/
- Wendy: are we looking at all vocabularies or just some? And are we looking at just DataCite-specific vocabularies or pulling in values from elsewhere?
- Joe: for National Gallery - important to be able to define own description
- Joe: you don't have to do everything the same way right away - but ultimately would be good to
- Kristian: the external vocabulary service it does require maintaining an external service and relying on it being available. would we be considering something like a Dynamic Vocabulary Fields? as an alternate approach. Where Instead of hardcoding a list of allowable values into the schema itself, you can design it to reference an external list stored somewhere else in datacite.
- Kristian: simplest solution - pulling the list out of the schema - having it move at a faster revision pace than the schema itself.
- Ted: DataCite controlled/code lists cover a range of things - some don't change much (nameType), some change more frequently (resourceTypeGeneral, contributorType, relationType)
- Ted: Other - doesn't have free text for contributorType; and no Other relationType
- Kristian: which ones are more important? How do we prioritize? For resourceTypeGeneral - impacts whole schema - because it is what you can register DOIs for (needs guidance). Providers/contributors need to be alerted that we've made the change
- Kelly: resourceTypeGeneral - used in multiple places
- Joe: comes back to the notion of having a vocabulary that is filtered, so you could use resourceTypeGeneral subset in 10.a (resourceType) and more options in 12.f (relatedIdentifier)
- Ted: A set of lists from ISO Geographic Metadata:
https://wiki.esipfed.org/ISO_19115-3_Codelists
- Kristian: question about validation - whether we do server, github, another approach - how do we do validation? E.g. if a term is deprecated
- Joe: need to version controlled lists as well as schema. Need to lock in old vocabulary versions

- Ted: what's the situation where we would deprecate a term?
- Kristian: would need to declare controlled vocabulary versions in metadata deposits
- Joe: versioning controlled lists as a family - not individually. (fine to link to other controlled terms in this - DataCite would have its own representation of this)
- Wendy: has anyone been asking for multi-lingual resource types?
- Examples of larger vocabularies with process:
 - Getty AT
 - <https://rbms.info/vocabularies/>
 - Larger vocabularies would have a team
- Considerations for editorial process:
 - Domain experience
 - Still RFC process
 - Automating suggestion based on "Other" free text (resourceTypeGeneral, dateType - others would need added free text fields)
 - Could start by looking at corpus for resourceType free text
- Ted: we also have elements of existing lists (e.g. HasMetadata) that are used primarily by a small # of repositories
- Joe: should there be a process to retroactively change records if we add an "Other" term? Migration process?

Actions

- Thread for GH discussion of controlled lists approach
- Starting to draft workflow

May 4, 2023

Agenda and 08:00 UTC Notes

Attendees:

GitHub skills sharing

See [notes from previous meeting](#)

1. Questions to consider:
 - a. What are you wondering how to do?
 - b. What's working well?
 - c. What could be improved?
2. Topics we can cover:
 - a. Notifications
 - b. Tagging people with "@"
 - c. Categories
 - d. Labels
 - e. Formatting posts - Markdown, code blocks for xml/json

- f. Threading
- g. Updating your profile

April 13, 2023

Agenda

1. What's next for the DataCite Metadata Schema (5.0)
2. Schema 4.5:
 - a. [Dublin core mapping update](#)
 - b. XSD and examples
3. Schema 5.0:
 - a. [Part 1 discussions](#)
 - b. Next step: formation of subgroups for Part 2 and Part 3

08:00 UTC Notes

Canceled

16:00 UTC Notes

Attendees: Anne, Kirsten, Ted, Wendy, Kelly, Alex

[Recording was made for those unable to attend. Contact Kelly.]

Notes: Anne

1. What's next for the DataCite Metadata Schema (5.0):
 - ▣ What's next for the DataCite Metadata Schema [Metadata WG]

Kelly shared a presentation for framing the 5.0 planning:

- The schema doc lists (in various places) the primary purposes of the metadata are citation, retrieval, and discovery. This will be explored further.
- The discipline-agnostic nature of the metadata is also emphasized. This text has been fairly constant since schema 3.x.
- Considerations for metadata should include both the creators and the consumers (APIs, aggregators, etc.), as well as the source of the metadata transmitted to DataCite (curators, data managers, etc.) and the end-user applications for the metadata.
- See presentation for tables summarizing needs and pain points for providers and consumers, current state of the system, trade-offs between convenience for providers and usability for consumers.
- Brief discussion regarding the slide showing the completion percentages for some optional fields and the difficulties associated with making sure providers understand the larger significance of fields not important in their context; the additional effort to obtain some optional fields; the lag time associated with

adopting new schema versions with new capabilities; and lack of resources to backfill metadata for existing DOIs when fields are introduced.

- Note that subgroups will be active working on specific aspects of 5.0 design and development.

Further discussion

- Question to Anne from Wendy re:how well the approach aligns with her personal “vision”. She still advocates formal information modeling, which would be a major change but potentially provides a number of changes supporting long-term management and usability.
- Ted notes that “discovery” is a fairly broad concept, and attempting to support broad-ranging “discovery” through the DataCite schema is problematic. It might be better to lean more heavily on citation. Kirsten notes that there are some fields in the schema that seem directly related to discovery on a few axes (like subject keywords, or fields with controlled vocabularies). Encouraging repositories to think about discovery when formulating metadata would improve the content. Kelly notes there are some platforms already harvesting from text fields for discovery purposes. Anne suggests that tagging schema properties with the area they particularly address (citation, reuse, discovery, etc.) may help guide creators in selecting and filling metadata fields. Kirsten notes that any schema changes.
- Jan - I’m just coming to this after my leave: In the bibliographic world we have a set of ‘User Tasks’: Briefly, they are Find, Identify, Select, and Obtain. ‘Find’ involves meeting a user’s search criteria through various attributes of an entity e.g. author/title. ‘Identify’ enables a user to confirm they have found what they looked for, distinguishing among similar resources e.g. version/language etc. ‘Select’ involves meeting a user’s requirements with respect to content, physical format, etc. or to reject an entity that doesn’t meet the user’s needs. ‘Obtain’ enables a user to acquire an entity through purchase, loan, etc., or electronic remote access, including rights. This might also help us categorise various metadata elements. To a certain extent the citation involves the user tasks of Identify and Select for example.

2. Schema 4.5:

- a. [Dublin core mapping update](#)
- b. XSD and examples

3. Schema 5.0:

- a. [Part 1 discussions](#)
- b. Next step: formation of subgroups for Part 2 and Part 3

March 30, 2023

Agenda and 16:00 UTC Notes

Attendees: Anne, Mary, Mike, Joe, Kirsten, Wendy, Kelly

3. Questions to consider:
 - a. What are you wondering how to do?
 - i. Finding the repository
 1. Search
 2. Star it - my stars
 3. Type "datacite" and look at dropdown
 4. Via link: <https://github.com/datacite/metadata-wg>
 - ii. You can reply via email to discussions
 - b. What's working well?
 - c. What could be improved?
4. Topics we can cover:
 - a. Notifications
 - i. "Watch"
 - b. Tagging people with "@"
 - i. Do we need a pinned issue with people's usernames/names?
 - ii. @ and then type name or handle
 - iii. List of watchers: <https://github.com/datacite/metadata-wg/watchers>
 - c. Categories
 - d. Labels
 - e. Formatting posts - Markdown, code blocks for xml/json
 - i. Markdown cheat sheet:
<https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>
 - ii. Official guide: <https://docs.github.com/en/get-started/writing-on-github>
 - iii. For XML formatting:
 1. ```xml
 2. (insert xml)
 3. ```
 - f. Threading
 - g. Flow charts:
 - i. Joe left an example, pulled from:
<https://research.ng-london.org.uk/modelling/>
 - h. Updating your profile
 - i. Can also edit profile to include name, add picture
5. Actions:
 - a. Add yourself to "Watch" the repository
 - b. Kelly to upgrade all to "Triage" level
 - c. Issues

- i. Look into Milestones?

March 9, 2023

Agenda

1. Schema 4.5: [DataCite to Dublin Core Mapping update](#)
2. Schema 4.5: [XSD and examples](#): for offline discussion
 - a. Deadline: March 31
 - b. Implementation work to start Q2
3. [Language property](#)
4. [Temporal coverage](#)
5. GitHub skill sharing
 - a. What are you wondering how to do?
 - b. Doodle poll for separate meeting upcoming

08:00 UTC Notes

Attendees: Isabel, Matthias, Anu, Vanessa

1. Schema 4.5
 - a. A mapping between Datacite 4.4 and DC was developed 2 years ago. Mapping 4.5 is a draft (work in progress) and we need feedback from WG members.
 - b. Anu: There is dc-datacite mapping developed by ARDC, see https://docs.google.com/document/d/1Fvw6KKV2900TsuKnvNbonzR4xoN_o5AzbxNZgRBXEQY/edit
 - c. Isabel: If you know any repositories that implement DCMI terms, please share with the WG. We can use their implementation to refine our mapping (**Action: WG members**)
 - d. Please share any existing work on mapping from dublin core to datacite schema (**Action: WG members**)
 - i. Matthias - dspace plugin might implement this mapping, see <https://wiki.lyrasis.org/display/DSDOC6x/DOI+Digital+Object+Identifier#OIDigitalObjectIdentifier-Metadadataconversion>
 - Also see DC-datacite (old) mappings <https://github.com/DSpace/DSpace/blob/main/dspace/config/crosswalks/DIM2DataCite.xsl>
 - https://github.com/caltechlibrary/epxml_to_datacite
 - ii. Vanessa - EPrints has also mappings (DC, DataCite etc.)
 - https://github.com/caltechlibrary/epxml_to_datacite
2. Isabel: plan to release version 4.5 before summer (no definite date yet).
3. Language property
 - a. Isabel: there are no guidelines on representing language in the datacite schema.

- b. Matthias - not sure if there is a single registry that covers all indigenous languages. For indigenous Australian languages see <https://collection.aiatsis.gov.au/austlang/search>
 - c. Anu - another source <https://glottolog.org/> (covering indigenous languages); a good summary of language tags ;) - <https://www.w3.org/International/articles/language-tags/>
4. Temporal coverage
- a. Anu
 - i. proposal follows iso 19115 (called 'extent' in later schema).
 - ii. parent element 'coverage' with sub elements spatial and temporal +1 Matthias & Vanessa
 - b. How can we specify the 'validity' period of a contributor, project, etc.?
 - i. Matthias - RAID led by ARDC is looking into this issue (work in draft).
 - ii. Vanessa: Would adding a (trusted) timestamp be a solution to this?
5. Github skill sharing - training will be organized as a separate meeting (not as part of the monthly meeting). Kelly to send a doodle invite soon.
6. Isabel - April meeting is on (no conflict with Easter break ;))

16:00 UTC Notes

Notes: Mike

Attendees: Ted, Jessica, Kirsten, Felix, Mike, Wendy, Alex

1. Schema 4.5: [DataCite to Dublin Core Mapping update](#)
 - a. Isabel and Jan have worked on update
 - b. Original mapping was just two columns (DataCite to DC)
 - c. New mapping: include DC simple, qualified, and DataCite local DC extension
 - d. Deadline for feedback: March 31
2. Schema 4.5: [XSD and examples](#): for offline discussion
 - a. Deadline: March 31
 - b. Implementation work to start Q2 (April)
 - i. Updates to Fabrica forms and REST APIs, etc.
 - c. Check: XSD file, "include" and 'examples' directories
 - i. Try to see where we can break it
 - ii. Especially: places where validation is failing (i.e., cannot enter accepted value or can enter unacceptable value)
3. General v4.5 discussion:
 - a. Use of Ringgold as an identifier type: it is fairly closed, paid service; should we take it out as an example type of identifier?
 - i. Not to forbid them, just to remove as an example in documentation
 - ii. Other examples in list may not be completely open (ISNI?)
 - iii. DataCite also includes many local identifiers and ARKs (sort of local)
 - iv. Will weigh in with Jan and Isabel to determine process for such revisions
4. [Language property](#)
 - a. Discussion has raised questions about what codes to use (e.g., ISO 693-1 or 693-3)

- i. Current recommendations are for values from IETF BCP 47, ISO 639-1
 1. BCP 47 provides guidance on how to utilize tags from different registries (and include region or script extensions to language tags for greater specificity—e.g., fr-CA for French as used in Canada)
 2. BCP 47 recommends keeping language tags as short as possible
 - ii. Need to be open for other options:
 1. Requests for support of indigenous languages
 2. Also interest in 693-3 [Need to note that schema can support this: make clear]
 - b. Group also discussed language attribute for metadata fields in XML and how this could be implemented in JSON version of schema
 - i. Consistency - where does lang show up in JSON
 - c. Also discussed titles in multiple languages: may be appropriate for v5 schema work
5. [Temporal coverage](#)
 - a. Group looked into actual occurrence of geographic locations in DataCite records (in about 11% of total DOIs)
 - i. Do not have stats on geoLocationPoints/Polygons
 - b. Linking temporal coverage to specific spatial coverage may be important (e.g., data refer to this location at this time)
 - c. Purpose of geospatial: applications for spatial search
 - d. Actual use—esp. With searchability—should factor into decisions
6. GitHub skill sharing
 - a. What are you wondering how to do?
 - b. Doodle poll for separate meeting upcoming

February 9, 2023

08:00 UTC

Attendees: Mohamed, Jessica, Vanessa, Matthias. Anusuriya, Isabel, Madeleine

16:00 UTC

Attendees: Wendy, Anne, Jan, Kelly, Mike, Ted, Alex, Joe

Regrets: Felix,

- Intros in GitHub: [Introductions · Discussion #4 · datacite/metadata-wg \(github.com\)](#)
- Version 4.5 will be released (approx. 3 months) after the (small) pending issues are resolved: e.g. Update of examples (in XML)
 - Anne can help with reviewing XML examples / validation against XSD
 - Kelly to look into migrating examples repo issues to main metadata-wg issues
- Overview of next steps

- Work is starting on version 5.0; check the Metadata Schema [Trello Board](#) for topics to be discussed in the WG. Schema release will approximately take 2 years.
- First changes to address:
 - [Creator/Contributor roles](#);
 - Temporal coverage; topic has been discussed before in the WG (probably for v. 4.4).
 - Language property;
- Full scope to be finalized in coordination with DataCite engineering team
 - Q: will this include refactoring? What does refactoring mean?
- Reminder of workflow: <https://schema.datacite.org/contribute.html>
 - Trello board could use more publicizing
- Discussion: Creator and Contributor roles proposal (<https://github.com/datacite/metadata-wg/discussions/7>)
 Add any comments you have on [GitHub](#). See also Schema change proposal document (link on GitHub) for background information.
 Matthias: in the definition of Creator, replace 'researchers' to 'people or organizations'. Creators are not only researchers, can be organization(s) too.
 When introducing controlled vocabulary for Creator/Contributor roles, we need to be aligned with DataCite's overall approach to controlled vocabularies.
- Future:
 - Workshop on how to use GitHub discussions
 - Data model call - brainstorming; should be somewhere in the upcoming weeks.
 - Creating citation from DataCite metadata - overview of current processes, content negotiation, CSL

January 12th, 2023

Attendees: Anne, Kelly, Kirsten, Madeleine, Jessica, Mohamed, Ted, Joe, Wendy, Alex, Vanessa, Jan, Mike

Regrets: Isabel

1. Welcome new members
 - a. Intros
 - b. Logistics
 - i. Google Drive: [Drive](#)
2. Next steps for 4.5:
 - a. Review final version: https://datacite-metadata-schema.readthedocs.io/en/4.5_draft
 - b. Comment by Wednesday, January 25
 - c. Examples to be finalized after schema
 - a. DC mappings need update - Isabel, Jan
 - b. XSD - DataCite staff

- i. Anne can support with validating examples against the XSD in March
- ii. Jsn schema possible for 5.0 - needs technical support.

3. Collaboration - <https://docs.github.com/en/discussions/quickstart>

- Can arrange github training

<https://schema.datacite.org/contribute.html>

December 15th 2022

Attendees: Wendy, Kelly, Ted, Mohamed, Isabel, Mike, Kristian

Regrets: Anne, Madeleine, Vanessa, Joseph

Note taking rotation list: Mike, Mohamed, Ted, Vanessa, Wendy, Kelly, Anne, Felix, Joseph, Madeleine

1. New WG members joining in January:
 1. Jessica Parland-von Essen, CSC - IT Center for Science (Finland)
 2. Matthias Liffers, Australian Research Data Commons
 3. Anusuriya Devaraju, Commonwealth Scientific and Industrial Research Organisation (CSIRO) (Australia)
 4. Alexander Whelan, New York University
 5. Kirsten Elger, GFZ German Research Centre for Geosciences
2. Outstanding issues - deadline for written comments: January 4
 - a. Instruments: <https://github.com/datacite/schema-docs/pull/24>
 - b. Distribution: <https://github.com/datacite/schema-docs/pull/23>
 - c. RelatedItem: <https://github.com/datacite/schema-docs/pull/25>
 - d. Other changes: <https://github.com/datacite/schema-docs/pull/26>
 1. Includes: Minor formatting issues, missing "other" dataType in appendix, reference to 'default' nameType; updates for relationTypes that are 'recommended for discovery'
3. Update from Examples sub-group
 1. Core examples for 4.5 will be:
 1. Examples with RelatedItem
 2. Example showing multilingual metadata
 1. regarding the multilingual XML example, this new guidance may be of interest to have a look at <https://www.coar-repositories.org/files/Implementation-Guidance-COAR-Multilingualism-1.pdf>
 3. Example for Dataset ResourceType
 4. Example for Instrument ResourceType
 5. Example with Distribution
 6. Full XML example (all properties)

2. Many of the 4.4 examples were based on existing DOIs. We have evaluated these and will reference the actual DOIs (where appropriate), to ensure the source is clear and reduce the need for maintenance from the Metadata WG. We hope to build a “bank” of real examples over time.
3. Work in progress; to be completed January.
4. Plan for release and outstanding actions
 - c. DC mappings need update - Isabel, Jan
 - d. XSD - DataCite staff
5. AOB

Bios:

JPvE has a background in digital humanities and has been involved in Open Science and digitalisation and has been working with research data management services at CSC in Finland for 8 years. Lately she has been involved in EOSC projects (FAIRsFAIR and FAIR-IMPACT) and is also a member of the ePIC consortium management board and coordinator of PID Forum Finland.

With a background in information science and information technology, Matthias leads the work on the Australian Research Data Commons PIDs roadmap, has experience in minting and managing identifiers for a range of outputs and is a member of the ROR curation board.

As repository manager for GFZ Data Services, the geosciences domain repository hosted at GFZ (Potsdam, Germany), one of my aims is to support data discovery via rich metadata, including with PIDs and linked-data vocabularies used in the geosciences domain. We are publishing data from the long-tail to large international networks in geodesy and geophysics and even beyond. In addition, I can contribute with my experiences on sample metadata as member of the IGSN DataCite Partnership Steering Group and long-year activities as IGSN Allocating Agent.

Alex Whelan is the Metadata Strategist at New York University Libraries. He brings a commitment to open access publishing and digital non-print collections as well as a deep interest in linked data. Alex's priority at NYU is optimizing the architecture for accessibility of online electronic resources, and he looks forward to bringing his knowledge of cooperative cataloging and national metadata standards to the ongoing challenge of DOI interoperability.

Anusuriya has extensive experience working with various national and international environmental research infrastructures and data initiatives, specializing in information modelling and developing workflows and services to support continued preservation and access to digital assets.

November 10 2022

Attendees: Isabel, Kelly, Mike, Madeleine, Anne, Joe, Felix, Mohamed, Wendy, Ted, Vanessa
Regrets: Jan

Note taking rotation list: Mike, Mohamed, Samantha, Ted, Vanessa, Wendy, Kelly, Anne, Felix, Joseph, Madeleine

1. RFC feedback overview: 📄 v4.5 RFC feedback: Nov 10
 - a. Summary documents in [v4.5 RFC Feedback](#)
 - b. Sourced from: [DataCite Metadata Schema 4.5: Request for Comments](#) and [schema-docs GitHub issues](#)
 - c. Input deadlines from WG members:
 - i. Instruments: Thu Nov 17
 - ii. Distribution: Thu Nov 17
 - iii. RelatedItem: Tue Nov 22
 - d. Notes:
 - i. Feedback gathered Sep. 12 - Oct. 31
 - ii. Biggest issues:
 1. Instruments
 - a. Key questions:
 - i. ResourceTypeGeneral: MeasuringInstrument vs Instrument
 - ii. Clarifications of relationTypes and definition/scope of instrument (as well as description guidance)
 - b. Intent behind 'instrument' may be fluid: specific one or generic 'type'? Hardware? Software? (Would have different roles/types)
 - i. Standardization may be challenging; are there controlled vocabularies that can be used for categorization? May require more research
 - c. DOI originally intended for a unique, specialized instrument
 - d. Proposed change to revise definition to refer to specific physical instance (and not general category)
 - e. Reference to "scientific purposes" may introduce confusion (is it scientific method or research domain?); may be better

to refer to 'research purpose'? Could be helpful for definition to help limit the scope of 'instruments' (i.e., it's not *anything* used in research). At the same time, DataCite philosophy is to let users adopt fields for their purposes

- f. Acquiring DOIs for instruments: responsibility of the instrument owner/manager (i.e., of a lab) as opposed to the researcher who used it?

g. Metadata WG members should add comments to Instrument document to information small group discussion

2. Distribution property--key questions in feedback included:

- a. Repeating distributions for multiple files
 - i. Based on DCAT 'distribution'
- b. mediaType declaration (no option specific to BagIt bag)
- c. byteSize: units and property name
- d. accessRights: disambiguate from "rights"
 - i. Related issue: entanglement of copyright/licensing information (which are different from access)
 - ii. Definition refers to 'access status'--might be more accurate field name (but is it used elsewhere in DataCite)
 - iii. COAR Access Rights vocab is more about the availability of content (e.g., metadata-only, public, embargoed)
- e. Relation to 'FAIR' digital object? We may want to take a deeper dive into this for 5.0

iii. Minor changes on

- 1. relatedItem,
 - a. Issue with backwards compatibility around recommended changes to isPublishedIn: address in next major version
- 2. publisher identifiers
 - a. People loved it!
 - b. Include example with RORs
 - c. Consider examples asynchronously
- 3. pre-registration/registration reports
 - a. Needs guidance on how to apply the Publisher field (may be addressed as part of broader Pub. guidance in 5.0)

iv. Other comments:

- 1. Minor formatting issues, DC mappings, missing "other" dateType in appendix, reference to 'default' nameType; updates for relationTypes that are 'recommended for discovery'

- e. Subgroup meetings are coming up: review feedback in preparation

2. Updated schema feedback process

- a. Refreshed the schema website to have a new page—Contribute—for documentation explaining the process:
 - <https://schema.stage.datacite.org/contribute.html>. (This is our staging site)
 - i. Includes calls to action ("Submit idea") and explains the stages of the process.
 - ii. Stages have been renamed for clarity - to avoid the implication that the ideas in what was called "Under Consideration" are the only ideas we are considering:
 - 1. Under Consideration → Community Discussion
 - 2. In Discussion → Metadata Working Group Discussion
 - iii. Reorganized website - with a menu bar and separate pages.
 - b. Switch from Productboard roadmap to Trello
 - i. Users will not have the "How important is this to you?" question to give private feedback on issues that are undergoing public discussion - instead we will link to PID Forum threads (or other appropriate place).
 - ii. Trello board is linked from the schema website
 - iii. Could balance use of Trello between public tracking and WG conversations
 - iv. This could also be helpful to get feedback on individual pieces of 5.0 (as opposed to comments on schema in its entirety)
 - c. Questions about how to increase community engagement:
 - i. Metadata Interest Group?
 - ii. Other products to allow conversation? <https://zulip.com/for/research/>
 - iii. Goal is to be more responsive with feedback/suggestions
3. AOB
- a. Plan for the rest of the year
 - i. Next meeting on Dec. 1
 - b. Call for nominations (Metadata WG, CESG, and STSG) open through November 14: see [Call for nominations: join a DataCite steering or working group!](#)

October 13 2022

Attendees: Kelly, Isabel, Wendy, Anne, Mohamed, Madeleine, Jan, Felix, Mike, Ted, Vanessa

Note taking rotation list: Kristian, Madeleine, Mike, Mohamed, Samantha, Ted, Vanessa, Wendy, Kelly, Anne, Felix, Joseph

- 1. Call for new members to go out soon
 - a. [Metadata WG rotation preferences](#) (background document: https://datacite.org/documents/Metadata_WG_ToR.pdf)
 - b. Currently have 12 members (excluding DataCite staff); ToR specifies 10-16
 - c. Clarify if there are open seats as well as representative on the STSG
- 2. XML Examples sub-group
 - a. Meeting scheduled - October 20 6am PDT

Will there also be worked on JSON examples? Not there at the moment.

3. Process for gathering schema ideas: [Schema Feedback Process](#)

Kelly is presenting the slides about the process.

Ted: it's not possible now to indicate on the Roadmap that the idea is not considered important (negative feedback).

Madeleine: new ideas will first be evaluated by DataCite staff before it will be added to the Roadmap. Submitters are not aware of that.

Anne: prioritize the schema data/information model to make it more sustainable for years to come; change control system that is more transparent. Should we keep expanding the schema as we do now?

4. Overview of RFC feedback to date: [v4.5 RFC feedback: Oct 12](#)

5. AOB

Meeting invites

September 8 2022

Attendees: Jan, Joe, Anne, Kelly, Wendy, Isabel, Madeleine, Mike, Mohamed, Ted

Note taking rotation list: Joseph, Kristian, Madeleine, Mike, Mohamed, Samantha, Ted, Vanessa, Wendy, Kelly, Anne, Felix

Documentation and community feedback process

1. **Draft Metadata Schema v4.5 - Request for Comments** [10 minutes]

[DataCite Metadata Schema 4.5: Request for Comments](#)

- a. Summary of feedback, focusing on the *Overview* section
- b. Approval to circulate the RFC to the community on September 12. Details of the process are here: [DataCite Metadata Schema 4.5: Community Feedback Process](#)

Notes: Not too many comments or questions internally. They have been responded to though. A few outstanding issues related to the current areas of discussion such as IGSN.

A few items will be updated such as links. Feedback process considered - not many issues raised with this. Perhaps we need a glossary of initials :-). Outreach responsibilities are being allocated please look. Need to **define** a more formal process for feedback. - Tested here plans to go forward in 5.0.

2. **"Future Directions" statements** in the *Request for Comments* document. [10 minutes]

- a. *Contributor roles* - Summary of feedback
- b. *Contributor roles* - Decision on approval
- c. *Related Item* - Summary of feedback
- d. *Related Item* - Decision on approval

Notes: Information to help people see the direction of future work. Areas listed are things that are actively being worked on, but were not ready for 4.5 or represent bigger changes thus are being considered for 5.0. Defining a more PID based structure for Contributor roles and expanding the scope of how related items are used in 5.0. Concerns raised about IGSNs but this will be covered later. These can also be seen as previews and an aim to promote further feedback and discussions, providing a wider reach/dissemination for ongoing work.

Related item plans are deliberately open as there is no current commitment to a direction, just an understanding that developments are required. This could lead to more structural changes so flagging the need for discussions and the gathering of use cases are very important.

Are related items used a lot: do we have stats here (MY) - Not yet but this data should be available by the end of the year. (Developments of the API underway)

Is it possible to remove things that are not used? (MY) This would be a breaking change so can only be done during a major version change.

We need to reconsider how developments are discussed - there may well need to be some major changes in the future to improve the efficiency and consistency of developments in the future. Key to aim for consistency of approaches across the schema Building on and developing rather than changing things too much - needs an overall approach rather than considering individual issues. Examining patterns for solutions etc.

Things can be complex without being complicated (AR).

Should the schema really solve everything or just the core things with the ability to connect well to other specializations.

It might be worth considering examining the standard methods for structuring models ... Poss Nov Dec (AR)

3. **Readthedocs** draft of metadata schema 4.5

[10 minutes]

https://datacite-metadata-schema.readthedocs.io/en/4.5_draft/

- a. Summary of feedback (Comments on the RFC and emails)
- b. GitHub issues: Discussion on how WG members should respond to ensure that we do not offer solutions, which may not be feasible for implementation in 4.5.

Notes: New work/approach prepared by (KS). Is the plan to allow people to comment on docs going forward. This is a possible plan. (KS) When we get feedback we need to offer solutions after discussion rather than adding the first thing springing to mind. Is there an approach to providing feedback for feedback Not ideal so far. Do not want to add too many tools here but some documentation of the feedback process. Ideally needs a Feedback Editor to do this well Currently a number of different ways to submit "issues" so it is harder to keep track of. There used to be a rota to keep track of this but currently it is not flowing. This is being discussed

within DataCite but it has not defined a specific solution yet. There are issues beyond just the Metadata group etc.

Transparency would be good. There is a roadmap discussion which could be more clearly disseminated. Even internally (AR) is not sure what comments have been made in the past.

4. XML Examples Sub Group

[10 minutes]

- a. Request for members to work on revising XML examples
- b. Proposal here: [DataCite Metadata Schema 4.5: Examples Strategy](#)

Notes: They are looking for people to do this - is there a deadline (AR) Nov time This will be worked on quite soon. Updating existing ones and potentially gathering new ones. AR offered to validate them. Commented on potentially modeling some of the examples (<https://research.ng-london.org.uk/modelling/>) (AR has a full dummy example which uses all aspects of the 4.0 schema, it will need to be updated for 4.5.

TH and AR discussed some format complications here.

Datacite XML - to JSON being considered - needs some non standard modeling.

5. IGSN [IGNS Proposal](#)

[10 minutes]

- a. Update on the review: The proposal document has been updated to only include the physical object definition change and the renaming of IGSN to International Generic Sample Number
- b. Decision on approval

Notes: The proposal for 4.5 changes are included but there are more discussions about further changes and considerations for 5.0.

A solution to identify DOI IGSNs as IGSNs is not yet in the metadata - this workflow is still being documented. The changes required to do it well will need to wait until 5.0

Should we include some comments on the future needs of IGSNs in the future plans section? (IB) - (KS) has suggested that this is possible and will try to add in something simple.

RelatedIdentifier does not really work as it breaks the scope description to some extent in most situations (MY) - (TH) this is generally a people issue rather than a technical one - the brand IGSN is important and needs to still exist even though they are going to be modeled as DOI type PIDs. This is an issue of DOIs families - types of DOIs - etc.

There is also a discussions about how the people solution actually breaks some of the technical scopes ... types is the way to go here (TH)

As a 5.0 question or potentially beyond - how “in general” should “extra” metadata coming from other groups of PIDs be absorbed into DOIs - this needs to have a standard process of connecting “extra” metadata which is not part of the core DataCite model.

Provenance information may also be a key issue here (AR) ... this is a general issue, but it is being pushed by the IGSN family. Identifier types, resource types etc still to be resolved.

6. StudyRegistration [StudyRegistration Proposal](#) [10 minutes]

- a. Update on the revised definition: The sub-group has revised the proposed definition of StudyRegistration to the following:

A detailed, time-stamped description of the research plan, often openly shared in a registry or published in a journal before the study is conducted to lend accountability and transparency in the hypothesis generating and testing process.

- b. Decision on approval

Notes: Simple definition change. Few issues or comments. Suggested changing “the” to “a” research plan. (JP) - (MY) asked about tools to create these - templates do exist but a tool is not known at the moment. Too specific to publishers. (OSF template). Provenance of this one inferred by examples.

AOB

Notes: No AOB issues were discussed.

[Metadata WG Members - ORCID and ROR IDs](#)

August 30 2022: IGSN proposal call


Attendees: Kelly, Cody, Matt, Joe, Mohamed, Isabel, Felix, Anne, Wendy,

1. How to reference IGSN IDs registered as DOIs through DataCite services
 - a. This is important because it may impact the relatedIdentifierType="IGSN" example
 - b. Example: an institution registers IGSN IDs for samples, and wants to reference these from a dataset. Should the relatedIdentifierType be IGSN or DOI?
 - c. Under the partnership agreement: IGSNs are DOIs. identifierType=DOI
 - d. Legacy IGSNs will be aliased to a DOI - handled by CNRI
 - e. Would it make sense to push this to 5.0, when we can address how to distinguish IGSN IDs?
 - f. MY: Answer should be DOI if the identifierType is DOI, relatedIdentifierType should be DOI. Could there be an attribute
 - g. JP: Flag for IGSN should not be in relatedIdentifier

- h. AR: Identifier is the primary identifier for the metadata, but it doesn't have any attributes
 - i. MY: What about attributes for re-registered Crossref DOIs?
 - j. AR: What is the goal in making sure that people know IGSN DOIs came from the IGSN in some sense? If the goal is specific set of metadata; aggregators to be able to discover legacy IGSNs
 - k. JP: identifierType other than DOI could cause problems for software that is relying on this
 - l. AR: least disruptive solution is an optional top-level element to the schema that can be ignored, restructure in 5.0
 - m. JP: metadata family for IGSNs?
 - n. AR: think of this as distinguishing the metadata, rather than the metadata itself
 - o. CR: there is no alternateIdentifier for new IGSNs; would just be the DOI.
 - p. MB: this is an entire community shifting and joining DataCite, as members - different from the periodicals case.
 - q. MB: We should stick with what it is (it is a DOI). for related identifier: legacy IGSNs can continue to be referenced as IGSN with the original IGSN. for new IGSNs, should be DOI.
 - r. MB: should there be an optional property now that we can use for 5.0?
 - s. KS: not sure the benefit given that we can distinguish by prefix / repository account (some institutions will register other DOIs, but under different repository accounts)
 - t. JP: could the contributor field be used? E.g. contributorType="Other", value = IGSN or the specific repository? Could be pre-populated for legacy ones
 - u. AR: Developing implementation options for IGSN clients as stopgaps for 4.5, and share with the IGSN community?
 - v. JP: Could it be in Subject? Single-element taxonomy?
 - w. MB: it is not necessarily an identifier type, but an identifier type or brand - specific to a subclass of physical objects
 - x. **Decision:** No recommendation to use relatedIdentifierType=IGSN for 4.5 due to the inconsistency with identifierType - these are DOIs.
2. Finalize the definition of PhysicalObject
 - a. A revised version is proposed in [Support for IGSN IDs](#) to remove the 3D aspect and to clarify that digital representations of physical objects should use other types. The language is adapted from DCMI, as suggested.
 - b. **Action:** All to review and leave remaining comments, final decision to be made Sept 8.
 3. If time: share background on aliasing process and legacy IGSN ID metadata
 - a. This is specific to the IGSN partnership agreement, it will be covered in support documentation and not form part of schema 4.5

August 25 2022

Attendees: Jan, Kelly, Isabel, Vanessa, Mohamed, Cody, Joseph, Felix, Wendy
Regrets: Anne, Madeleine, Sarah, Mike

1. Welcome Cody!
2. IGSN changes, [including definition change for PhysicalObject](#) (10 min)
 - a.  Support for IGSN IDs
 - b. Physical Object definition:
 - i. Dimensionality and time - is it necessarily 3D?
 1. Dates can be used to convey range
 - ii. Sometimes when developing ontologies, we define things based on how it is used, not because it needs to be limited
 - iii. Not removing “three-dimensional” at this point simply for the sake of making it more open; revisit in 5.0.
 - iv. “Features-of-interest” in allowed values refers to places where samples are taken from
 - v. There are 2D samples too—physics context (thin films, membranes)
 - vi. Would a researcher not put a sample in because they thought it wasn’t 3D?
 - vii. What about “analog” (vs. “digital”) as opposed to 3D?
 - viii. **Decision:** Revise definition to change “three dimensional” to “physical”. Add note to clarify that digital representations should use the other resource types.
 - c. RelatedIdentifier:
 - i. Clarification from Cody: re-registered (legacy) IGSN IDs will have an alternateIdentifier with the existing handle **and** a relatedIdentifier with relationType IsIdenticalTo. This is part of the IGSN crosswalk recommendation.
 - ii. What is the difference between alternateIdentifier and relatedIdentifier with IsIdenticalTo? IsIdenticalTo is used for the same resource located in a different place.
 1. Ted: This may not have been considered by the IGSN-DataCite WG. Needs to be reconciled (either modify IsIdenticalTo or modify the recommendation)
 2. Wendy: IsIdentical is for cases like double posting in Zenodo/IR, important to have a type for the location
 - iii. What happens to IGSNs that have been previously cited?
 - iv. Existing IGSNs will be aliased and continue to resolve to new DOIs.
 - v. How will users know which DOIs are IGSN IDs to use the IGSN relatedIdentifierType when referencing it?
 - vi. **Decision:** Schedule a separate call to discuss - Kelly to send Doodle poll.
3. Rotating notetaker: Felix

- a. Rotation list: Joseph, Kristian, Madeleine, Mike, Mohamed, Samantha, Sarah, Ted, Vanessa, Wendy, Kelly, Anne, Felix
- 4. Schema change proposal sub-group updates (15 min)
 - a. [StudyRegistration resourcetypeGeneral - 4.5 DataCite Metadata Change](#)
 - i. Typically used for medicine, but expanding to other fields (e.g. psychology)
 - ii. Discussion of potential overlap with OutputManagementPlan
 - iii. Should there be a scope note indicating which fields it is typically used in, or originates from?
 - iv. It is starting to cross domain boundaries
 - v. **Decision:** Sub-group to revisit description to emphasize the hypothesis piece; needs to be sufficiently distinguished from general project proposal/project plans.
 - b. ContentURL: see [21. Distribution](#)
- 5. Next steps for 4.5 (35 min)
 - a. Updates from documentation format sub-group
 - i. Revised structure with new sections for Mappings and Guidance; proofreading in progress. Draft online here: https://datacite-metadata-schema.readthedocs.io/en/4.5_draft/
 - b. Community feedback process
 - i. Planned draft release date: **Monday, September 12**
 - 1. Initial deadline: Monday, October 17
 - 2. Extended deadline: Monday, October 31
 - ii. [DataCite Metadata Schema 4.5: Request for Comments](#)
 - 1. Public Google doc for gathering comments
 - 2. Links out to Readthedocs site so full schema can be reviewed
 - 3. Please review before September 2nd**
 - iii. Documentation of process for DataCite staff and Metadata WG, including outreach strategy:
 - [DataCite Metadata Schema 4.5: Community Feedback Process](#)
 - c. Proposed section in RFC on “future directions”
 - i. Outlining current thinking on Contributor role and Related Item
 - ii. Sharing with the community that these suggestions are not lost, but they involve breaking changes and are best suited for 5.0
 - iii. See section of RFC: [Future Directions: Version 5.0 and beyond](#)
 - iv. **Actions:** Review RFC and draft on Readthedocs
- 6. AOB:
 - a. Please confirm your ORCID iD and your institution’s ROR ID here:
 - [Metadata WG Members - ORCID and ROR IDs](#)
 - i. RORs might be offset (resolved)
 - b. Calendar invites
- 7. Next call:
 - a. Revising XML examples - see proposal:
 - [DataCite Metadata Schema 4.5: Examples Strategy](#)

July 14 2022

Attendees: Jan, Kelly, Kristian, Wendy, Ted, Anne, Mike, Sarah, Isabel

Regrets: Mohamed, Joseph, Felix, Madeleine

1. Rotating notetaker: Anne
 - a. Rotation list: Felix, Joseph, Kristian, Madeleine, Mike, Mohamed, Samantha, Sarah, Ted, Vanessa, Wendy, Kelly, Anne
2. **ContentURL:** There are no new updates from the Sub-Group at this point. They are still working on the feedback provided in the previous session.
 - a. Still working on incorporating feedback received to date.
3. **RelatedItem:** The Sub-Group has a change proposal for 4.5
 - a. [RelatedItem - 4.5 DataCite Metadata Change](#) . (10 min) Presented by Kelly (lead)
 - b. Group has generated ideas to improve the guidance for using the current implementation (see above document).
 - i. Clarifying consistencies
 - ii. Removing ambiguities
 - iii. Examples (working examples)
 - c. RelatedItem should be revisited for schema 5.0.
 - d. Of the two use cases, the ability to provide the container details for a resource seems to have the greater community interest and numbers of users waiting. The second use case is still of interest, but to a smaller number of users (at least as documented).

Interesting to note that use case 1 (container) was not discussed during the development of this property (TH).
 - e. Documentation changes to property changes are relatively minor. They are noted in red in the document (strike-through indicates proposed deletions).
 - f. Questions:
 - i. Next Steps? We should consider timelines, targeting September, perhaps.
 - ii. Schema 5.0 changes? The WG noted that relatedItem is trying to solve too many problems at once, while also leaving some gaps (a chapter in a book in a series, for example, is not supported). The goal is to better support traditional literature. There might be a better way to do that, but it will probably require non-backwards compatible changes to the schema.
 - iii. Was container information really a part of the discussion? Some disagreement among memories, but there are examples from prior presentations, so it was involved at some level.
 - iv. Are there relationships like the resource:relatedItem for other elements of the schema like, for example, the funding stream (an agency has a funding program which is administered through an institutions, etc.)? Probably, but should that be encoded into DataCite's metadata? Brief

discussion about the implications of expanding and maintaining metadata, automated PID linking, and related topics. These are considerations that need to be looked at carefully for Version 5.0

4. **Next steps for schema 4.5** development and validation (30 min)
 - a. [Schema 4.5 Release Plan](#)
 - b. Timeline for release with schema 4.5 in March 2023. Kristian notes that expecting a full month's development in December is probably not realistic. Kelly explains that the timeline is aimed primarily at Working Group activities, which would be light in December.
 - c. Kelly has been investigating ReadTheDocs as a new format for the Schema document. This produces both an online version and also provides options to generate PDF (and ePub) formats of the document from the same source. The backend is a GitHub repository. Comments can be enabled on the pages (at our specification), and generate issues on the back end.
 - d. The Release Plan includes a list of groups to review specific changes and how to reach them to solicit feedback.
 - e. The process of requesting feedback is tricky to manage to get good, actionable feedback. The CrossRef solicitation document from their 2020 schema update was presented an example. Proposal to use a combination of Google Docs and ReadTheDocs comments. Discussion:
 - i. What did/does the Library of Congress do in collecting feedback?
 - ii. Formal feedback requires some investment of time to read, understand, and formulate.
 - iii. It is common to see review deadlines of about 6 weeks, then extended by a week or two.
 - iv. Will there be enough time for discussion once reviews are in?
 - v. Is it possible to get some feedback from the DataCite developers to head off major ambiguities in documentation that delay full implementation?
 - vi. Should DataCite be looking to create a Beta Test group to exercise schemas, tools, and documents before formal release?
 - vii. A broad example set is useful for both users and developers. In particular, for the developers it would be useful to carry the examples through to what might displayed on a landing page as e.g., a reference. Complex examples are also important to include.
5. Sub-group to discuss documentation format (10 min)
 - a. Proposal to create this. Volunteer by emailing Kelly and Jan.
 - b. The 4.5 document will provide a practical test and chance to exercise the ReadTheDocs format.
6. AOB
 - a. Reminder to join <https://groups.google.com/g/datacite-metadata>

June 9th 2022

Attendees: Kristian, Kelly, Sarala, Isabel, Madeleine, Wendy, Anne, Joe, Mohamed, Vanessa
Regrets: Jan, Ted, Sarah, Mike

1. Rotating notetaker
 - a. Today's notetaker: Kelly
 - b. Rotation list: Felix, Anne, Joseph, Kristian, Madeleine, Mike, Mohamed, Samantha, Sarah, Ted, Vanessa, Wendy, Kelly
 - c. Flexible to re-order as conflicts arise

2. [Metadata Schema support for content URLs](#) (Kristian) - **20 min**
 - a. Mike and Sarala overview of the proposal
 - b. Feedback collection on the Miro [board](#)
 - c. Problem to be addressed: there is no standardized way to retrieve *resources* (files) identified with DOIs. Currently, reliance on landing pages or different media APIs, which are inconsistent across repositories.
 - d. Use cases: direct access to data for researchers, harvesters, integrators, etc.; better compatibility with other standards, such as DCAT; support data retrievability using standard communication protocols.
 - e. Implementation uses subset of properties from DCAT
 - f. Who will provide this metadata?
 - i. Biggest contributors already provide most of this information on landing pages - Zenodo, Dataverse, Figshare, Pangea, Dryad...
 - g. What if the dataset consists of hundreds of files?
 - i. Use a package file - bagit, zip, tar, etc.
 - h. Is the intention of the proposal to describe dataset file granularity?
 - i. No - should be used to give machine-access to collection, not to describe a collection

3. Align Contributor roles with CRediT list (Jan / Kelly) - **20 min**
 - a. Please review the proposal document below and decide which option you would prefer the WG to take.
 - i. Proposal: [Align ContributorType with CRediT ontology](#)
 - b. Option 1. Include the CRediT values into the current DataCite contributorType list.
 - c. Option 2. Delay until v5.0 when we can carry out a major review of the contributorType values and any new sub-element and attributes.

Discussion:

A preliminary survey among the group yielded the two options below. A further survey was taken to decide between the two. A merged controlled list of both was not a comfortable idea for the majority, and the subgroup is leaning toward Option 2. Is that reasonable?

Question about the amount of overlap/confusion between the two lists. Ted provided a crosswalk that shows correspondences, but note that the mappings are not 1:1 in either direction. There is also an issue with people who have multiple roles, all of which should be reflected. (Contributor role is currently single-valued.)

Question whether the alignment with CRediT makes sense for DataCite. The interoperability aspect is a strong argument, but the lack of alignment between the two systems complicates that both going forward and for interpreting legacy metadata.

If this is delayed until 5.0, the subgroup should probably begin developing the corresponding documentation to support discussions and validation with the membership for the eventual implementation.

4. Request for feedback via email:
 - a. [Publisher identifiers subgroup](#) (Isabel)
 - i. Confirming schema change proposal
 - ii. Originally discussed possibility of adding more than 1 publisher and making identifier subproperty repeatable
 1. Current recommendation - temporary solution for 4.5: use Publisher property for main publisher; include secondary publisher in Contributor property
 2. For 5.0, looking at making this property repeatable
 3. Need to clarify difference between Publisher and Contributor, and which roles to use for “secondary” publishers or other entities contributing to publishing.
 - iii. Recommendation to get identifiers for repositories, publishers, etc. where possible (Wikidata, re3data)
 - iv. The sub-group drafted guidelines on how to understand who the publisher is, depending on the resource type - this is currently paused until we decide on format for extra user guidelines in parallel to the schema specification.
 - v. **Action:** Group to provide feedback on the above recommendations via email.
5. New sub-groups
 - a. relatedItem documentation
 - b. Pre-registration resourceTypeGeneral
6. Licensing for schema documentation
 - a. Proposal to use CC-BY for 4.5 release

May 12th 2022

Attendees: Isabel, Kelly, Wendy, Mohamed, Anne, Ted, Mike, Sarala, Joe, Sarah, Vanessa, Felix
Regrets: Jan, Kristian, Madeleine

Timeline for schema release (Kelly) - **5min**

- a. May-June 2022: Decisions on current proposal items.
- b. August 2022: Complete draft schema proposal.
- c. September 2022: Community validation phase. Share draft proposal for feedback from consortium leads, regional expert groups, service providers, and those who initiated the proposed change.
- d. October 2022: MWG completes revisions based on community feedback.
- e. November 2022: Final schema documentation completed for handoff to development team (work is scheduled for Q4).

[Support for PIDs for instruments](#) and [Appendix](#) - confirm decisions from April meeting (TBA) - **10min**

- f. Add 'Instrument' to the controlled list values of resourceTypeGeneral
- g. Add relationType pair 'IsUsedBy' / 'Uses' to the controlled list values of the relationType property to describe the relationship between a research activity or output and the instrument.
- h. Use descriptionType 'TechnicalInfo' to include detailed information associated with an instrument instance.

[Publisher identifiers subgroup](#) (Isabel) - **15min**

- i. We would like to get feedback as regards:
 - i. Overview of potential identifiers for Publishers: is the list complete? What do you think it is missing?
 - ii. Our proposal to leave the definition of Publisher as it is in the schema documentation but facilitate its correct usage through additional guidelines and examples through different use cases: do you agree? If so, what is the best way to go about extra guidelines for users?
 - iii. Our proposal to include new subproperties in the schema to indicate type of Publisher identifier, schema and ID: do you agree?

[Metadata Schema support for content URLs](#) (Kristian) - **15min** → **next meeting**

- a. The Sub Group invites the WG members to keep adding ideas/solutions to the [brainstorming section of the MIRO board](#).

AOB → **Kelly to email group**

- j. Pre-registration resource type proposal
- k. Documentation for relatedItem

April 14th 2022

Attendees: Ted, Kelly, Sarah, Jan, Anne, Mike, Madeleine, Helena, Wendy, Mohamed

1. Agenda
 - a. Focus on 3 items
 - b. Other subgroup items ready for feedback:
 - i. [Add ROR to the Publisher](#)
2. [Support for PIDs for instruments](#) (20min)
 - a. Lead: Madeleine
 - b. Notes:
 - i. Need a longer term plan for how DataCite will develop the schema, including controlled lists. Adding Instrument as a resourceTypeGeneral works as a stopgap solution.
 1. **Decision:** Add Instrument to ResourceTypeGeneral.
 - ii. HasUsedIn / WasUsedIn, MeasuredBy / Measures
 1. Other use cases beyond instruments?
 2. Should it be WasUsedBy? Or Uses / IsUsedIn? Collects / IsCollectedBy? Observes / IsObservedBy?
 3. **Decision:** Add one relationType pair: Uses / IsUsedIn or IsUsedBy
3. [CreatorType for Creators](#) (20min)
 - a. Lead:?
 - b. Notes:
 - i. Idea to get rid of Creator, and add types Creator and Publisher to contributorType?
 - ii. Longer term idea - backwards compatibility would be a problem. Consider for later versions.
 - iii. Right now, can't properly credit editors - problem for tenure review. Original use case from change proposal - see https://wiki.tib.eu/confluence/download/attachments/86261208/DataCiteAuthor_en.pdf?version=1&modificationDate=1566563185000&api=v2
 - iv. Are there other cases where creator would need additional types - i.e. who is the creator of an instrument (manufacturer)?
 - v. **Action:** Further subgroup discussion needed. Possible to push this to 5.0 if a larger change is needed.
4. [Align Contributor roles with CRediT list](#) (20min)
 - a. Lead: Jan
 - b. Could WG members please look at the "Questions" in the comments for the Categorise Solutions section of the document.
 - c. Notes:
 - i. **Action:** WG members to respond to Jan's questions via email (Jan to send email). Subgroup will collect responses and review the original use cases/proposal.

Next meeting:

5. Overview of updated Working list
<https://wiki.tib.eu/confluence/display/datacite/Working+list+for+Version+4.5+or+5.0>
(Kristian to go through updates on Distribution/Maturity indicators subgroups and recent decisions taken)
6. [Add ROR to the Publisher](#)
 - a. Overview of potential identifiers for Publishers
 - b. Our proposal to leave the definition of Publisher as it is in the schema documentation but facilitate its correct usage through additional guidelines and examples through different use cases
 - c. Our proposal to include new subproperties in the schema to indicate type of Publisher identifier, schema and ID
7. [Metadata Schema support for content URLs](#)
 - a. Lead: Kristian
 - b. The Sub Group invites the WG members to keep adding ideas/solutions to the [brainstorming section of the MIRO board](#). We will be reviewing them in the following weeks.
8. Processing community enquiries and proposals

March 10th 2022

1. Welcome and introduction
2. Review of queries and change proposals from the public groups [Recurring item]
 - [Process for reviewing DataCite metadata schema queries and change proposals](#).
 - [Review of queries and change proposals to date](#).
3. [Metadata Schema support for content URLs](#)
 - a. Lead: Kristian G
 - b. Participants:
 - i. Vanessa Gabriel
 - ii. Mike Shallcross
 - iii. Sarala Wimalaratne
 - iv. Jan Ashton
 - c. [Sub-group notes](#)
 - d.
4. Data maturity/quality indicators (Isabel, Samantha and Marleen)
<https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=164761319?>
This sub group is on hold at the moment - participants may wish to take part in another sub group for the time being.

5. [Support for PIDs for instruments](#) (Madeleine)

6. [Add ROR to the Publisher](#) (Isabel)

7. [CreatorType for Creators.](#)

e. Lead: ?

f. Participants:

i. Vanessa Gabriel

Note: useful to liaise with the *Align contributor roles with CRediT list* Sub Group.

8. [Align Contributor roles with CRediT list](#) (Jan)

A.O.B

February 10th 2022

1. New validated suggestions from Datacite membership. They can be assigned for WG discussion

- [Metadata Schema support for content URLs](#)
 - **Lead:** Kristian G
 - Participants:
 - (Vanessa Gabriel)
 - Availability for Subgroup ContentUrl discussion:
https://doodle.com/poll/gf3hmm4thrxdnn89?utm_source=poll&utm_medium=link

Discussion

Need to discuss if we want to provide access to the data directly via URL from the metadata. The focus of DataCite to date has been for the DOI pointing to a Landing page.

- [CreatorType for Creators.](#)
 - **Lead:** ??
 - Participants:
 - Vanessa Gabriel

Discussion

Need to be careful not to conflate Creator with contributorType. However, it is useful to liaise with the *Align contributor roles with CRediT list* Sub Group.

2. Working list Sub Group updates (ca. 25 minutes)

- [Align Contributor roles with CRediT list](#) (Jan)

Discussion

A strategy could be to point domain specific vocabularies externally from the schema. We probably need to review all of the controlled lists and have a policy on our strategy for them.

Possible to have generic top level lists of values which could be the default or mandatory in order to provide for cross-domain search and discovery etc. Alongside a lower level optional set of values or pointers to other vocabularies to provide domain specific values, without overloading the schema itself. Allow for local values as well as published vocabularies.

Possible downside of this approach could be the proliferation of crosswalks required due to the multiple value instances.

- [Support for PIDs for instruments](#) (Madeleine)

Discussion

Comments on the scope of the work - what comprises an 'instrument' and its components etc.

- [Add ROR to the Publisher](#) (Isabel)

Discussion

Do we need to revisit the definition of 'Publisher'? There is confusion and some overlapping of the terms between it and contributorType e.g. distributor; hosting institution. If we change the definition we must be aware of backwards compatibility.

AOB

- [Using RelatedItem metadata to generate a formatted reference](#)
 - We still need help to clarify the use of relatedItem implementation.
 - Could we [schedule](#) a call?

January 13th 2022

1. Welcome and introduction to new members! (ca. 15 minutes)
 - a. Vanessa Gabriel from the University Library of the LMU Munich
 - b. Joseph Padfield from the National Gallery, London
 - c. Sarah Ramdeen from Columbia University
 - d. Wendy Robertson from the University of Iowa
 - e. Mike Shallcross from the University of Michigan, Inter-University Consortium for Political and Social Research
2. Working list Sub Group updates (ca. 25 minutes)

- a. [Align Contributor roles with CRediT list](#) (Jan)
- b. [Add ROR to the Publishe](#)
- c. https://gbr01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.google.com%2Fdocument%2Fd%2F1REPI82D_x3p66YoXreEG58k-6QUr2hMs4Qn1z0QONM%2Fedit%23heading%3Dh.cded8372egm4&data=04%7C01%7C%7C4cff4efe97648827ad708d9b8e5c608%7C21a44cb7f9c34f009afabd1e8e88bcd9%7C0%7C0%7C637744120327611911%7CUnknown%7CTWFpbGZsb3d8eyJWljoiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6Ikk1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=X%2FHQIVZAWiWZPdtwedvXEKA3KcU%2FCYtdRsk032%2Fe7y0%3D&reserved=0r element (Isabel)
- d. [Support for PIDs for instrumen](#)
https://gbr01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.google.com%2Fdocument%2Fd%2F1REPI82D_x3p66YoXreEG58k-6QUr2hMs4Qn1z0QONM%2Fedit%23heading%3Dh.cded8372egm4&data=04%7C01%7C%7C4cff4efe97648827ad708d9b8e5c608%7C21a44cb7f9c34f009afabd1e8e88bcd9%7C0%7C0%7C637744120327611911%7CUnknown%7CTWFpbGZsb3d8eyJWljoiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6Ikk1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=X%2FHQIVZAWiWZPdtwedvXEKA3KcU%2FCYtdRsk032%2Fe7y0%3D&reserved=0ts (Madeleine)

3. Overview of WG processes and DataCite Metadata Schema Change Proposals Workflow (ca. 10 minutes)

- a. .New requests for amendments to the metadata schema should be added to the DataCite Product Board. Members can submit a 'New idea' at <https://portal.productboard.com/datacite/1-roadmap-2021/tabs/1-under-consideration/su bmit-idea>
- b. The DataCite Product team will aim to provide feedback within a week. They will add either a new Roadmap card or link to an existing card if the proposal has been raised previously.
- c. The Product team will contact the submitter for more information and feedback.
- d. The submitter will present their proposal at the next Open Hours meeting, inviting feedback and/or support from other members who can add their comments to the relevant card.
- e. Once a card has been created members can add feedback, comments and support for the proposal.
- f. After the Open Hours meeting, the Product team will send the proposal to the Metadata Working group along with feedback from the Open Hours meeting. This is so that the WG have the context and background information about the proposal.
- g. Based on the outcomes of Open Hours and discussions with the submitter and the broader community of users, the Product Engineering team may validate the proposal. They will then ask the Metadata WG to work on including it into the metadata schema.

- h. The Metadata WG will update or add the change proposal to the Working List on the Wiki for the next release of the metadata schema.
 - i. The WG will ask for interested parties to join/lead a Sub Group to work on the proposal.
4. Confirm process for notification by DataCite staff of any metadata issues on the Google Public discussion list and PID Forum (DataCite Chat Room) etc. (ca, 10 minutes)
 - a. e.g. email added to the Google discussion list about date of collection.
 5. AOB
 - a. How to format a reference using **RelatedItem**?
 - i. Is it intended that the relatedItem information is used to generate a reference only when the type is **IsPublishedIn**?
 - ii. If there is more than one title in relatedItem, which xtitle should be used?
 - b. Update
 - i. We will be sending feedback on the items below on the 24th week:
 1. [Add 'creatorType' to the Creator element](#)
 2. [Metadata Schema support for content URLs](#)
 - ii. We will be validating the items below in the next OpenHours on the 26th
 1. preregistrations
 2. Create separate schemas or 'profiles' for specific resource types
 3. Data Maturity / Quality Indicators RelationType

Action items

- Add the new members to the tibwiki - <https://wiki.tib.eu/confluence/display/datacite/Working+list+for+Version+4.5+or+5.0>

December 9th 2021

Attendees:

Ted Habermann, Metadata Game Changers

Announcement: Liz will be replacing Sarala within the group from next year

1. Update on DataCite Open Hours: any new topic has been validated, other updates? (ca. 5 minute discussion) (Sarala/Liz)
 - a. [Roles for creators: Enlargement DataCite Schema with CreatorType](#)
 - b. [Support for content URLs](#)
2. [Working list Sub Groups](#) (ca. 40 minutes discussion)
 - a. [Align Contributor roles with CRediT list](#) (Jan)
 - b. [Add ROR to the Publisher element](#) (Isabel)

Action from November meeting: Sarala to find out “Add ROR (not affiliationIdentifier?) to the Publisher element” requires structural changes to the schema xml or json

- c. [Support for PIDs for instruments](#) (Madeleine)
- d. [Data Maturity / Quality Indicators](#) (Marleen)
3. Management of regular review of metadata issues on the Google Public discussion list and PID Forum.
4. AOB
5. Date of next meeting: Thursday 13th January 2022

November 11th 2021

1. New member applicants (ca. 20 minutes discussion)

[Call for applications - Metadata Working Group \(Responses\)](#)

What expertise do the applicants offer that fill gaps or that we would like in the WG? ([See the summary document](#))

- Decision on how many applicants to invite.
- Decision on which applicants to invite.
- Decision: Invite top 5 voted applicants
 - **(7) Sarah Ramdeen**
 - Resource types – physical objects (scientific samples and specimens).
 - **(5) Vanessa Gabriel**
 - Documentation - Best Practice Guide; DataCite metadata generator and use cases.
 - **(4) Wendy Robertson**
 - Resource types – Journal publishing and Digital images
 - Interoperability - Crossref
 - **(3) Mike Shallcross**
 - Schema extension – metadata enhancement and augmentation
 - Technical – Python scripts
 - Interoperability – metadata sharing and reuse.
 - **(3) Joseph Padfield**
 - Resource types – Heritage science data
 - Interoperability – in relation to FAIR
- Action:
 - Co-chairs to email the new members in December
 - On board new members in January

2. Effective use of tools (ca. 5 minutes review)

Wiki and Google Docs results of online vote.

- Decision on which method to use.
- Decision
 - Google docs for working documents
 - Wiki for final documents
- Action - Marleen to check with KIT about installing addons for the wiki

3. Working list Sub Groups (ca. 15 minutes discussion)

- Confirm that we have enough members willing to lead and work on each of the three **validated** Change Proposals.
- Review interest in any other 'under consideration' proposals.
- Plan work of Sub Groups and approval by the main Working Group e.g.

each sub-group to have someone willing to take the lead on overseeing input to the relevant Template 'form' and report back to the main meeting on progress?

- Actions
 - Jan will discuss the leads for the sub groups offline
 - Sarala to find out "Add ROR (not affiliationIdentifier?) to the Publisher element" requires structural changes to the schema xml or json

4. Action points in addition to the working list: (ca. 15 minutes discussion)

- Good practice documentation with examples

- Information about metadata completeness

- Regular review of metadata issues on the Google Public discussion list and PID Forum.

- Regular and proactive communication: Google Public discussion list; PID Forum's DataCite Chat Room and Blog posts.

- Suggestions for any members who would like to work on any of these issues. Similar to a Sub Group.
- Liaison with DataCite communications team.

5. AOB

Oct 14, 2021

Agenda

- Account of the WG presentation in September's members meeting (Ted).
 - Presentations went very well
 - [IB] Maybe we can identify what we want to improve following the poll feedback. Eg: improvements to documentation, DataCite services etc.
 - [TH] following the polls, affiliation should not be mandatory
 - [MY] It would be good to know who attended the session. Is it the DOI minters or researchers who searches using Commons
 - [AR] If the DOI minters are further away from publications, affiliation is not useful. For example: instruments
 - [TH] How can we help members understand the benefits of some of the properties such as people, organizations, funders etc.
 - [IB, AR] Todo: Improve documentation, examples, good practice for different resource types
 - [AR] May be a quick overview of how complete the metadata per repository to encourage users to add more metadata
- Any other relevant feedback from the members meeting (Sarala)
 - The 3 sessions worked quite well and good participation overall
 - Any thoughts for improvements for next time
- Community feedback to the Metadata WG (Kristian).
 - Currently the schema.datacite.org website suggest to visitors to contact the WG via the Public Google group?:

Feedback

If you have any comments on

- any challenges you face with the current schema - i.e. what's not working for you?
- any ideas you have for additions that would make the schema work better for you, or
- questions you have about the schema,

please contact us via our [Google discussion group](#). We will review the suggestions and comments on a monthly basis, and reply using this same discussion forum.

- However the Public Google Group (<https://groups.google.com/g/datacite-metadata>) has rarely been updated with responses from the Working Group in the last years.
- Is the Public Google Group not being used anymore by the WG? Should we change the advice on the schema website? Shall we advise to put the suggestions only in the DataCite Roadmap?
 - Disadvantages:
 - No open discussion of suggestions: Suggestions to the roadmap remain private until DataCite manually publishes them.

- Questions cannot be captured. Separated channels for questions, challenges, and/or suggestions will need to be set in place.
 - [KS] It is very quiet on the google groups so people might assume that the group is not active.
 - [IB] Many ways to contact the group which is hard to manage
 - [JA] Everything is funneled through the roadmap.
 - [AR] General community can communicate via the public google group, PID forum, blog etc. Members can put it forward through the roadmap.
 - [JA] Metadata group need to be active by posting in the groups. Use a set of blog posts to drive discussions. Maybe once a month. Eg: Eventdata, subject classification blog posts.
 - [SW] Document the process on the website.
 - Google drive for the Metadata WG for the meeting minutes and working documents
 - https://drive.google.com/drive/folders/0AAg89a_ncNzsUk9PVA
 -
 - [JA] Figure out the best way forwa
 - DataCite Proposed Metadata Change Template rd between wiki/google templates
 - Discussion around the 3 items already validated and reminder for WG members to join any of them (Isabel/Jan).
 - [Align Contributor roles with CRediT list](#)
 - [Add affiliationIdentifier to the Publisher element](#) >>>> replaced by [Add ROR \(not affiliationIdentifier?\) to the Publisher element](#)
 - [Support for PIDs for instruments](#)
 - Make Creator's affiliation Mandatory?
- Pages created for each of them at
https://drive.google.com/drive/folders/1-JnCS2w_XfEJJKac95-oy3EKmqqc_2kk
- Update on the call for applications for the WG (Sarala)
 - [Call for applications - Metadata Working Group \(Responses\)](#)
 - We have 11 applications
 - On the form we said we are going to onboard 3 but there is space up to 6
 - How do we move this forward

Sep 9, 2021

We have a new co-chair, welcome Jan!
 Kelly Stathis joined as a new member today, welcome!

1. Terms of Reference (Matt)

TOR is available at https://datacite.org/documents/Metadata_WG_ToR.pdf

Will be working/reporting to the Board and ED and no longer will be reporting to STSG.

2. WG member recruitment (Matt)

We have three applicants:

- David Milward, Metadataworks Limited
- Wendy Robertson, University of Iowa
- Martina Stockhause, German Climate Computing Center (DKRZ)

[AR] Circulate the call for applications at the member meeting.

[SW] Figure out who we want to onboard depending on what we want to achieve.

[SW] Timeline suggestion to onboard next year

[MB] Chairs can decide on the timelines and the path forward

3. Announcements (Isabel)

Sophie Roy will be stepping down at the end of September. She will be replaced by Kelly Stathis from the Canada Consortium

4. Update on preparations of participation in DataCite member meeting, Sept.22 (Ted)

Link to the slides

A subgroup of the metadata group met last week and Ted is working on the slides

5. Update of PE activities on the working list (Sarala)

PE team has validated - ready to be worked on

- [Align Contributor roles with CRediT list](#)

PE team has it under consideration Roadmap

- [Add affiliationIdentifier to the Publisher element](#)
- [Support for PIDs for instruments](#)
- [Enable reverse or reciprocal relationships for relatedIdentifiers of type DOI](#)***

For next open hours

- [Support for PIDs for conferences](#)
- [Support for machine access to download data directly](#)
 - [Support for checksums to better support FAIR principles](#)
 - Include checksum of the content - Will

PE have it under consideration for the future

- Copyright Information - Anne ***
- Links to PID persistence policies - ***
- [Editor Role Field](#) - Anne ***
- [Create separate schemas or 'profiles' for specific resource types](#)
- [Data Maturity / Quality Indicators](#)
- [Support for grant identifiers](#)
- [Support for new property "Distribution"](#)

6. Update on ongoing work DataCite application profile (Jan) [DATACITE DUBLIN CORE APPLICATION PROFILE.pptx](#)

7. Any other business (Isabel)

- For a closer alignment with FAIR Principles, these 2 guidelines may be useful to identify and evaluate all potential metadata improvements:

RDA FAIR Data Maturity Model. Specification and Guidelines

<https://zenodo.org/record/3909563#.YSdzyEtxdPZ>

FAIRsFAIR Data object Assessment Metrics,

<https://zenodo.org/record/4081213#.YSjER0ztaUk>