GREI Metadata and Search Subcommittee: Recommendations from DataCite schema version 4.4

Version 01: Last updated 2023-06-29

Overview

One goal of GREI is to support interoperability and discovery of datasets across repositories by establishing common metadata standards for the generalist repositories. Having focused on an agreed standard, the DataCite Metadata Schema 4.4, the GREI Metadata and Search subcommittee has set its Year 2 goal for repositories to build on their existing work on metadata for research datasets. Focusing on a few high-level use cases for data sharing and searching allowed the group to move forward to identify specific metadata beyond the DataCite required properties metadata that would meet the needs of those use cases.

With the inclusion of DataCite as a GREI stakeholder, more opportunities have been reviewed to now provide a recommendation to the GREI repositories to add additional metadata fields and enhance the quality of the metadata being provided. The subcommittee has continued having detailed discussions ensuring that GREI repositories collect and provide metadata in a way that is useful to all stakeholders.

With this in mind, the GREI Metadata and Search subcommittee has created this recommendation to strongly encourage that each repository member collect the following metadata to support the generalist repository use cases for sharing, discovering and tracking the impact of data.

We also hope this common metadata schema will be useful for data repositories beyond GREI to improve interoperability across data repositories and across the NIH data landscape.

Recommendation

The document lists strongly encouraged metadata to be collected by each GREI repository in alignment with the metadata collected by DataCite's optional metadata properties. Where applicable, the values and vocabularies that repositories are encouraged to use have also been reviewed by the subcommittee and included in the recommendations.

While these are the recommendations of the subcommittee to-date, the goal of the subcommittee is for the recommended common schema to evolve based on community feedback, new use cases, and emerging community initiatives to allow for updates and additions to the strongly recommended fields in the future.

For DataCite's property definitions, cardinality/quantity constraints, and other usage notes, see DataCite Metadata Schema Documentation v4.4.

ID	DataCite-Property	Subcommittee recommendations about property values
1	Identifier	Not applicable.
2	Creator	Not applicable.
2.1.a	nameType	Use controlled list values from DataCite: "Personal" or "Organizational".
2.2	givenName	Not applicable.
2.3	familyName	Not applicable.
2.4	nameIdentifier	For names of people, use ORCID IDs. For names of organizations, use ROR IDs.
2.4.a	nameIdentifierScheme	For names of people, use "ORCID". For names of organizations, use "ROR".
2.4.b	schemeURI	For names of people, use ORCID's URI, https://orcid.org/ For names of organizations, use ROR's URI, https://ror.org/
2.5	affiliation	Use names of organizations from the Research Organization Registry (ROR).
2.5.a	affiliationIdentifier	Use IDs from ROR in URL format (e.g., https://ror.org/05h1kgg64).
2.5.b	affiliationIdentifierScheme	Use "ROR".
2.5.c	schemeURI	Use ROR's URI, https://ror.org/
3	Title	Not applicable.
4	Publisher	Not applicable.
5	PublicationYear	Not applicable.
6	Subject	
6.a	subjectScheme	
6.b	schemeURI	
6.c	valueURI	
8	Date	Not applicable.

8.a	dateType	Use controlled list values from DataCite.
8.b	dateInformation	Not applicable.
10	ResourceType	
10.a	resourceTypeGeneral	Use controlled list values from DataCite.
12	RelatedIdentifier	
12.a	relatedIdentifierType	Use controlled list values from DataCite.
12.b	relationType	Use controlled list values from DataCite.
12.f	resourceTypeGeneral	Use controlled list values from DataCite.
15	Version	Not applicable.
16	Rights	Not applicable.
16.a	rightsURI	
16.b	rightsIdentifier	
16.c	rightsIdentifierScheme	
16.d	schemeURI	
17	Description	Not applicable.
17.a	descriptionType	For general descriptions of the deposit, select the descriptionType "Abstract" from DataCite's controlled list.
19	FundingReference	
19.1	funderName	
19.2	funderIdentifier	Use IDs from the CrossRef's Funder Registry or from ROR.
19.2.a	funderIdentifierType	Select "Crossref Funder ID" or "ROR" from DataCite's controlled list.
19.2.b	schemeURI	Use https://www.crossref.org/services/funder-registry/ or https://ror.org/
19.3	awardNumber	
19.3.a	awardURI	

19.4	awardTitle	

Feedback on Recommendations

The subcommittee plans to continue gathering feedback on the proposed recommendations. In particular the subcommittee will be asking the Use Case and Metrics subcommittees to review the recommendations as part of their activities and provide feedback about what is missing or should be changed.

In addition to the two subcommittees, the Metadata and Search subcommittee asks other participants of the GREI project to review the recommendations and give any feedback or commentary. The subcommittee asks that this feedback be given starting in June 2023 with a plan for all feedback to be delivered no later than June 07, 2023 from the GREI CWG.

Next Steps

Once feedback has been gathered on the proposed recommendations from the GREI CWG, this document will be updated and made public-facing and available for further feedback as a first iteration. By the start of July 2023, we aim to finalize the recommendations and publicize them through the activities of the Community Engagement subcommittee. Each GREI repository can begin reviewing them so that by December 2023, they can communicate their plans to follow the recommendations as part of their GREI-funded work in Years 2, 3, or beyond.