

DataCite Metadata Schema 4.4 to Schema.org Mapping

Documentation (1.0, Dec 2022)

Citation:

Stathis, K., Ross, C., Dreyer, B., & Vierkant, P. (2022). *DataCite Metadata Schema 4.4 to Schema.org Mapping*. DataCite. <https://doi.org/10.14454/3w3z-sa82>

Authors:

Kelly Stathis, <https://orcid.org/0000-0001-6133-4045>

Cody Ross, <https://orcid.org/0000-0002-4684-9769>

Britta Dreyer, <https://orcid.org/0000-0002-0687-5460>

Paul Vierkant, <https://orcid.org/0000-0003-4448-3844>

Contents

Goals	1
Background	2
Challenges we had creating the mapping	3
Mapping	5
Example	16

Goals

Building bridges to other domains

The DataCite Metadata Schema is a general, domain-agnostic metadata schema used for DataCite DOI registration. To improve interoperability, the DataCite Metadata Schema can be mapped, or crosswalked, to commonly used or domain-specific metadata standards.

This mapping from the DataCite Metadata Schema to Schema.org builds on existing efforts to produce crosswalks. For example, the DataCite Metadata Working Group has produced a mapping from DataCite to Dublin Core.¹ DataCite Content Negotiation also returns DataCite DOI metadata in various formats, including Schema.org, JATS, and BibTeX.² This mapping is based on the same mapping used by DataCite's metadata conversion library (bolognese)³, with modifications.

Search Engine Optimization

Schema.org is an initiative by search engines to add structured data markup to web pages widely implemented across the web. Google reads the structured data of the web content and uses it to display it on its results pages. Therefore, implementing Schema.org results in higher visibility of the content on the web. This accounts for research outputs on landing pages of repositories or academic journals too.

Researchers and research institutions use Schema.org as a way of search engine optimization (SEO) while improving the findability of their research outputs (e.g., through Google Dataset Search). One example of academic SEO to improve the findability aspect of the FAIR principles is Bioschemas.org, which represents profiles of Schema.org created by the life sciences community.⁴

¹ DataCite Metadata Working Group. (2021). DataCite to Dublin Core Mapping 4.4. DataCite e.V. <https://doi.org/10.14454/qn00-qx85>.

² <https://support.datacite.org/docs/datacite-content-resolver>

³ <https://github.com/datacite/bolognese>

⁴ <https://bioschemas.org>

Background

To foster the adoption of the DataCite Metadata Schema in the engineering sciences and other research domains, its mapping to Schema.org was one of the goals of the NFDI4Ing⁵ Seed Funds for the Base Services project. This mapping is meant to build bridges between the domain-agnostic DataCite Metadata Schema and specialized community standards while supporting the use of PIDs and metadata within NFDI4Ing as an essential component for the implementation of the FAIR principles for research data. Together with the Best Practice Guidelines for the DOI Registration of Research Outputs, this mapping shall enable researchers from engineering sciences to properly describe their research outputs and improve the findability of their research.

Challenges we had creating the mapping

The Schema.org type hierarchy

Schema.org⁶ has a type hierarchy. Hereby *Thing* is the most general type, with eleven immediate children—the most relevant being *CreativeWork*. Each type inherits properties from its parent and has its own properties. *CreativeWork* also has a long list of sub-properties, including *Dataset*, *Article*, and others relevant to research.

For this mapping, we chose to focus on the *CreativeWork* type because *CreativeWork* and its subtypes cover the majority of resource types that DataCite DOIs are used for. We also selected *CreativeWork* because the DataCite Metadata Schema—while increasingly used for a wide range of resources—is most suited to creative works, having originally been designed for use with datasets. Therefore, the Schema.org properties used below either come directly from *CreativeWork* or are inherited from *Thing*.

⁵ <https://nfdi4ing.de>

⁶ <https://schema.org>

Lack of equivalent properties

Not every property and sub-property in the DataCite Metadata Schema has an equivalent in Schema.org. While we were able to map all of the required properties, several of the optional (sub-)properties had to be omitted from the crosswalk.

Several of the DataCite properties have type attributes, which complicated the mapping further. For example, the *Date* property maps to different Schema.org properties (e.g. “dateCreated”, “dateModified”) depending on the *dateType* (e.g. “Created”, “Updated”). For some type values, we did not find a suitable equivalent in Schema.org (for example, there is no equivalent for a Date with *dateType* “Submitted”).

Of the properties that could not be perfectly mapped, *RelatedIdentifier* posed the biggest challenge. Several *relationTypes* did not have an equivalent in Schema.org that applied to the CreativeWork type, including the relationTypes for versioning (“IsPreviousVersionOf”, “IsNewVersionOf”, “HasVersion”, and “IsVersionOf”).⁷

Unfortunately, there was no generic equivalent to *RelatedIdentifier* that could be used regardless of relationType.⁸

Crosswalks depend on context

An optimal mapping from the DataCite Metadata Schema to Schema.org metadata would depend on how a given repository interprets the DataCite Metadata Schema generally, and may even vary from record to record. As such, while the mapping provided may serve as an initial starting point for implementation, repositories should review the Schema.org properties in detail and customize their mappings as needed.

⁷ While there are [predecessorOf](#) and [successorOf](#) properties, these can only be used with the type ProductModel.

⁸ The property [isRelatedTo](#) can only be applied with the types Product and Service.

Mapping

ID ⁹	DataCite Property	Schema.org Property	Additional Information
n/a	URL	url	While URL is not part of the DataCite Metadata Schema, it is required for DOI registration.
1	Identifier	@id	
2	Creator	author	
2.1	creatorName	author.name	
2.1.a	nameType	author.@type	For nameType "Personal", use @type value "Person". For nameType "Organizational", use @type value "Organization".
2.2	givenName	author.givenName	
2.3	familyName	author.familyName	

⁹ The top-level properties are indicated in bold with sub-properties below.

2.4	nameIdentifier	author.@id	
2.5	affiliation	author.affiliation.name	For author.affiliation.@type, use "Organization".
2.5.a	affiliationIdentifier	author.affiliation.@id	
3	Title	name	
4	Publisher	publisher.name	For publisher.@type, use "Organization".
5	PublicationYear	datePublished	
10	ResourceType	additionalType	
10.a	resourceTypeGeneral	@type	Value depends on resourceTypeGeneral:
			resourceTypeGeneral @type
			Audiovisual MediaObject
			Book Book
			BookChapter Chapter
			Collection Collection
			ComputationalNotebook SoftwareSourceCode

			ConferencePaper	Article
			ConferenceProceeding	Periodical
			DataPaper	Article
			Dataset	Dataset
			Dissertation	Thesis
			Event	-
			Image	ImageObject
			InteractiveResource	-
			Journal	Periodical
			JournalArticle	ScholarlyArticle
			Model	-
			OutputManagementPlan	-
			PeerReview	Review
			PhysicalObject	-
			Preprint	Article

			Report	Report
			Service	-
			Software	SoftwareSourceCode
			Sound	AudioObject
			Standard	-
			Text	ScholarlyArticle
			Workflow	-
			Other	-
6	Subject	keywords		
7	Contributor Mapped by contributorType: <ol style="list-style-type: none"> Editor Producer RightsHolder Sponsor <i>All other contributorTypes</i> 	[PROPERTY] varies: <ol style="list-style-type: none"> editor producer copyrightHolder sponsor contributor 		
7.1	contributorName	[PROPERTY].name		

7.1.a	nameType	[PROPERTY].@type	For nameType "Personal", use @type value "Person". For nameType "Organizational", use @type value "Organization".
7.2	givenName	[PROPERTY].givenName	
7.3	familyName	[PROPERTY].familyName	
7.4	nameIdentifier	[PROPERTY].@id	
7.5	affiliation	[PROPERTY].affiliation.name	For contributor.affiliation.@type, use "Organization".
7.5.a	affiliationIdentifier	[PROPERTY].affiliation.@id	
8	Date Mapped by dateType: 1. Copyrighted 2. Created 3. Issued 4. Updated 5. Withdrawn	Varies: 1. copyrightYear 2. dateCreated 3. datePublished 4. dateModified 5. expires	For dateType "Copyrighted", copyrightYear is YYYY only.
9	Language	inLanguage	
11	AlternateIdentifier	identifier.value	For identifier.@type, use "PropertyValue".

11.a	alternateIdentifierType	identifier.propertyID	
12	RelatedIdentifier	<i>variable</i>	The property for RelatedIdentifier varies based on relationType (12.b).
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> • Cites • References • IsSupplementedBy 	citation.@id	
12.f	resourceTypeGeneral	citation.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> • IsCitedBy • IsReferencedBy • IsSupplementTo 	@reverse.citation.@id	
12.f	resourceTypeGeneral	@reverse.citation.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> • HasPart 	hasPart.@id	

12.f	resourceTypeGeneral	hasPart.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> • IsPartOf • IsPublishedIn 	isPartOf.@id	
12.f	resourceTypeGeneral	isPartOf.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> • Describes • Documents • Reviews • IsMetadataFor 	about.@id	
12.f	resourceTypeGeneral	about.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> • IsDescribedBy • IsDocumentedBy 	subjectOf.@id	
12.f	resourceTypeGeneral	subjectOf.@type	Use the same mapping as 10.a resourceTypeGeneral.

12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> IsReviewedBy 	review.@id	
12.f	resourceTypeGeneral	review.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> HasMetadata 	subjectOf.@id	
12.c	relatedMetadataScheme	subjectOf.additionalType	
12.f	resourceTypeGeneral	subjectOf.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> IsDerivedFrom IsVariantFormOf 	isBasedOn.@id	
12.f	resourceTypeGeneral		Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> IsSourceOf IsOriginalFormOf 	@reverse.isBasedOn.@id	

12.f	resourceTypeGeneral	@reverse.isBasedOn.@type	Use the same mapping as 10.a resourceTypeGeneral.
12	RelatedIdentifier when relationType (12.b) is: <ul style="list-style-type: none"> IsIdenticalTo 	sameAs	Note: Because sameAs is a URL value, there is no @type.
13	Size	contentSize	
14	Format	encodingFormat	
15	Version		
16	Rights		From the Rights property (16), only the sub-property rightsURI is mapped.
16.a	rightsURI	license	
17	Description Mapped by descriptionType: <ol style="list-style-type: none"> Abstract All other descriptionTypes 	Varies: <ol style="list-style-type: none"> abstract description 	
18	GeoLocation	spatialCoverage	For spatialCoverage.@type, use "Place".
18.1	geoLocationPoint	spatialCoverage.geo	For spatialCoverage.geo.@type, use "GeoCoordinates".

18.1.1	pointLongitude	spatialCoverage.geo.longitude	
18.1.2	pointLatitude	spatialCoverage.geo.latitude	
18.2	geoLocationBox	spatialCoverage.geo.box	For spatialCoverage.geo.@type, use "GeoShape".
18.2.1-1 8.2.4	westBoundLongitude eastBoundLongitude southBoundLatitude northBoundLatitude	spatialCoverage.geo.box	The two points (bottom left and top right) should be concatenate and delimited by a space: <i>southBoundLatitude,westBoundLongitude northBoundLatitude,eastBoundLongitude</i>
18.3	geoLocationPlace	spatialCoverage.address	
18.4	geoLocationPolygon	spatialCoverage.geo.polygon	For spatialCoverage.geo.@type, use "GeoShape".
18.4.1 18.4.1.1 18.4.1.2	polygonPoint pointLongitude pointLatitude	spatialCoverage.geo.polygon	All polygonPoints should be concatenated and delimited by a space: <i>pointLatitude,pointLongitude pointLatitude,pointLongitude ...</i>

19	FundingReference	funder	For funder.@type, use "Organization".
19.1	funderName	funder.name	
19.2	funderIdentifier	funder.@id	
19.3	awardNumber	funding.identifier	For funding.@type, use "Grant".
19.3.a	awardURI	funding.@id	
19.4	awardTitle	funding.name	

Example

The following example shows a subset of the properties in the DataCite Metadata Schema.

DataCite XML Metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<resource
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="http://datacite.org/schema/kernel-4
http://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.82247/EGAQ-TQ60</identifier>
  <creators>
    <creator>
      <creatorName nameType="Personal">Stathis, Kelly</creatorName>
      <givenName>Kelly</givenName>
      <familyName>Stathis</familyName>
      <nameIdentifier nameIdentifierScheme="ORCID"
schemeURI="https://orcid.org">https://orcid.org/0000-0001-6133-4045</nameIdentifier>
      <affiliation affiliationIdentifier="https://ror.org/04wxnsj81" affiliationIdentifierScheme="ROR"
schemeURI="https://ror.org">DataCite</affiliation>
    </creator>
    <creator>
      <creatorName nameType="Personal">Ross, Cody</creatorName>
      <givenName>Cody</givenName>
      <familyName>Ross</familyName>
      <nameIdentifier nameIdentifierScheme="ORCID"
schemeURI="https://orcid.org">https://orcid.org/0000-0002-4684-9769</nameIdentifier>
      <affiliation affiliationIdentifier="https://ror.org/04wxnsj81" affiliationIdentifierScheme="ROR"
schemeURI="https://ror.org">DataCite</affiliation>
    </creator>
    <creator>
      <creatorName nameType="Personal">Vierkant, Paul</creatorName>
      <givenName>Paul</givenName>
      <familyName>Vierkant</familyName>
      <nameIdentifier nameIdentifierScheme="ORCID"
schemeURI="https://orcid.org">https://orcid.org/0000-0003-4448-3844</nameIdentifier>
      <affiliation affiliationIdentifier="https://ror.org/04wxnsj81" affiliationIdentifierScheme="ROR"
schemeURI="https://ror.org">DataCite</affiliation>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en">DataCite Metadata Schema 4.4 to schema.org mapping</title>
  </titles>
  <publisher>DataCite</publisher>
  <publicationYear>2023</publicationYear>
  <resourceType resourceTypeGeneral="Report"/>
  <dates>
    <date dateType="Created" dateInformation="2023">2023</date>
  </dates>
  <language>en</language>
  <relatedIdentifiers>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="Cites"
resourceTypeGeneral="Text">https://doi.org/10.14454/qn00-qx85</relatedIdentifier>
```

```

</relatedIdentifiers>
<sizes/>
<formats>
  <format>application/pdf</format>
</formats>
<version/>
<rightsList>
  <rights rightsURI="https://creativecommons.org/licenses/by/4.0/legalcode">Creative Commons Attribution 4.0
International</rights>
</rightsList>
<descriptions>
  <description xml:lang="en" descriptionType="Abstract">Metadata mapping from the DataCite Metadata Schema (4.4) to
Schema.org.</description>
</descriptions>
<fundingReferences>
  <fundingReference>
    <funderName>Nationale Forschungsdateninfrastruktur</funderName>
    <funderIdentifier funderIdentifierType="ROR">https://ror.org/05qj6w324</funderIdentifier>
    <awardNumber awardURI="https://nfdi4ing.de/de/nfdi4ing-seed-funds-2022/">
    <awardTitle>NFDI Seed Funds</awardTitle>
  </fundingReference>
</fundingReferences>
</resource>

```

Schema.org metadata

```
{
  "@context": "http://schema.org",
  "@type": "Report",
  "@id": "https://doi.org/10.82247/egaq-tq60",
  "url": "https://example.org/datacite-schemaorg",
  "name": "DataCite Metadata Schema 4.4 to schema.org mapping",
  "author": [
    {
      "name": "Stathis, Kelly",
      "givenName": "Kelly",
      "familyName": "Stathis",
      "affiliation": {
        "@type": "Organization",
        "@id": "https://ror.org/04wxnsj81",
        "name": "DataCite"
      },
      "@type": "Person",
      "@id": "https://orcid.org/0000-0001-6133-4045"
    },
    {
      "name": "Ross, Cody",
      "givenName": "Cody",
      "familyName": "Ross",
      "affiliation": {
        "@type": "Organization",
        "@id": "https://ror.org/04wxnsj81",
        "name": "DataCite"
      },
      "@type": "Person",
      "@id": "https://orcid.org/0000-0002-4684-9769"
    },
    {
      "name": "Vierkant, Paul",
      "givenName": "Paul",
      "familyName": "Vierkant",
      "affiliation": {
        "@type": "Organization",
        "@id": "https://ror.org/04wxnsj81",
        "name": "DataCite"
      },
      "@type": "Person",
      "@id": "https://orcid.org/0000-0003-4448-3844"
    }
  ]
}
```

```
],
  "abstract": "Metadata mapping from the DataCite Metadata Schema (4.4) to
Schema.org.",
  "license": "https://creativecommons.org/licenses/by/4.0/legalcode",
  "inLanguage": "en",
  "encodingFormat": "application/pdf",
  "dateCreated": "2023",
  "datePublished": "2023",
  "citation": {
    "@id": "https://doi.org/10.14454/qn00-qx85",
    "@type": "ScholarlyArticle"
  },
  "publisher": {
    "@type": "Organization",
    "name": "DataCite"
  },
  "funder": {
    "@id": "https://ror.org/05qj6w324",
    "@type": "Organization",
    "name": "Nationale Forschungsdateninfrastruktur"
  },
  "funding": {
    "@id": "https://nfdi4ing.de/de/nfdi4ing-seed-funds-2022/",
    "@type": "Grant",
    "name": "NFDI Seed Funds"
  }
}
```