

# DataCite Metadata Schema Changes: Request for Comments

Shared on: 26 March 2024

Feedback requested by: 6 May 2024

---

<b>Overview</b>	<b>2</b>
What we are looking for	2
What happens to your feedback	2
<b>Proposed Changes to DataCite Metadata Schema Properties</b>	<b>4</b>
Language property	4
Make Language repeatable and recommended	4
Clarify supported language codes	5
<b>Proposed Changes to Controlled Lists</b>	<b>7</b>
contributorType	7
New contributorType: Translator	7
relationType	7
New relationType: IsTranslationOf	7
New relationType: HasTranslation	7
dateType	8
New dateType: Coverage	8
relatedIdentifierType	8
New relatedIdentifierType: RRID	8
New relatedIdentifierType: CSTR	9
resourceTypeGeneral	9
New resourceTypeGeneral: Project	9

## Overview

The [DataCite Metadata Working Group](#) invites community feedback on proposed changes to the DataCite Metadata Schema. These changes are in response to suggestions from DataCite community members.

This RFC contains an assortment of changes that may be released in different versions—e.g., some changes may be released in Version 4.6, while others may be scheduled for later versions. The implementation details and release timing of specific changes will depend on complexity, backward compatibility, and the DataCite development roadmap.

This document will be open for public comment through **May 6, 2024**.

## What we are looking for

We are interested in your comments, questions, and suggestions related to the proposed changes in this document. We invite you to consider these three questions:

- Would the change work for your use case? Why or why not?
- Is there something you would like to see implemented differently?
- Which aspects of the change would benefit from clarification, if any?

You can leave comments and suggestions directly on this document. If you are reading this in PDF, you can access the Google Doc here:

[https://docs.google.com/document/d/1oaxTKV010BoRiBb\\_n0QY8mIFmhacUBvfk1LicYHnh0A](https://docs.google.com/document/d/1oaxTKV010BoRiBb_n0QY8mIFmhacUBvfk1LicYHnh0A)

Our goal is to make this feedback process as accessible and transparent as possible. If commenting on the Google Doc does not work for you, please send feedback to [support@datacite.org](mailto:support@datacite.org) and indicate whether you would like us to post a comment on your behalf. You can also email us with any questions about the process.

If you have additional suggestions for Metadata Schema changes that are not in this RFC, [please share them with us via our suggestions process](#).

## What happens to your feedback

DataCite staff and the Metadata Working Group will consider the feedback to help us iterate on the schema design. Please note that while we will actively monitor comments on this document, we may not reply directly to every comment.

Your feedback is appreciated and helps make the DataCite Metadata Schema better. Thank you!

## Proposed Changes to DataCite Metadata Schema Properties

The following are proposed changes to the DataCite Metadata Schema properties.

### Language property

#### Make Language repeatable and recommended

Resources may contain multiple languages, and may not always have a “primary” language. Currently, the DataCite Language property is non-repeatable, requiring users to select a single “primary” language of the resource even when a single resource has multiple languages.

To support multilingual resources (as in, resources that contain multiple languages), we propose to make the language property repeatable (0-n occurrences).

For monolingual resources that also exist in another language, we introduce new relationTypes “IsTranslationOf” and “HasTranslation” to establish links between translated versions. Each version should use the Language property to specify the languages present in that specific version exclusively. In other words, each unique resource identified by a DOI may possess a language property reflecting its language(s).

*Note: The change to make Language repeatable is breaking because the existing XML structure has a single <language> element. This would be implemented in a major version.*

In addition, we suggest changing the obligation level from *Optional* to *Recommended* to reflect the importance of the language property for indexing and searching. The DataCite Metadata Schema has three obligation levels:

- **Mandatory (M)** properties must be provided;
- **Recommended (R)** properties are optional, but strongly recommended for interoperability; and
- **Optional (O)** (but not specifically recommended) properties provide richer description.

### Obligation

Current obligation	Proposed obligation
Optional	Recommended

### Occurrence

Current occurrence	Proposed occurrence
0-1	0-n

### Definition

Current definition	Proposed definition
The primary language of the resource.	The <b>language(s)</b> of the resource.

### Clarify supported language codes

The [current documentation for the Language property](#) recommends values from IETF BCP 47 or ISO 639-1 language codes. Examples of 2-letter codes from ISO 639-1 are provided, but IETF BCP 47 is not explained.

Unlike ISO 639-1, [IETF BCP 47](#) is not a list of language codes. It contains RFC 5646 (“Tags for Identifying Languages”), which describes a *syntax* for language codes and the format of the IANA Language Subtag Registry. The IANA Language Subtag Registry is also represented on W3C’s page [Choosing a language tag](#), which explains its relationship with ISO codes:

*In the past, when dealing with lists of ISO codes, there were sometimes multiple codes for a given language - there could be a 2-letter code and one or two 3-letter codes. This ambiguity is resolved by the IANA Subtag Registry: only one code is listed per language. (If an ISO 2-letter code exists, that will be the code, otherwise it will be a three-letter code.) The registry maintainer also coordinates the ongoing evolution of the registry with developments in the ISO world.*

We propose to update the “Allowed values, examples, and other constraints” to direct users to the [IANA Language Subtag Registry](#) and include resources from W3C on selecting a language tag.

Because the ISO 639 standards are well recognized, we propose to maintain the existing reference to ISO 639-1 and add references to ISO 639-2 and ISO 639-3 to clarify that these code lists are supported.

We note that the [XML Schema language datatype](#) already supports codes following the syntax described in RFC 5646. As such, this change only impacts the Metadata Schema documentation and does not change what values are considered valid.

### Allowed values, examples, and other constraints

<i>Current "Allowed values, examples, and other constraints"</i>	<i>Proposed "Allowed values, examples, and other constraints"</i>
<p>Recommended values are taken from <a href="#">IETF BCP 47</a>, <a href="#">ISO 639-1 language codes</a>. Examples: en, de, fr</p>	<p>All values must follow the pattern defined by <a href="#">XML schema language type</a>.</p> <p>Recommended values are taken from the <a href="#">IANA Language Subtag Registry</a>. See the <a href="#">W3C guidance on Language tags in HTML and XML</a> and <a href="#">Choosing a Language Tag</a> for how to construct language tags using the IANA Language Subtag Registry, keeping the tag as short as possible.<sup>1</sup></p> <p>Examples constructed from the IANA Language Subtag Registry:</p> <p>language:</p> <ul style="list-style-type: none"> <li>● en</li> <li>● fr</li> <li>● mas</li> </ul> <p>language+extlang:</p> <ul style="list-style-type: none"> <li>● zh-cmn</li> <li>● sgn-ukl</li> </ul> <p>Alternatively, code lists such as <a href="#">ISO 639-1</a>, <a href="#">ISO 639-2</a>, <a href="#">ISO 639-3</a>, and other recognized sources<sup>2</sup> may be used directly.</p>

<sup>1</sup> From [Choosing a Language Tag](#):

“Language tag syntax is defined by the IETF’s BCP 47. In the past it was necessary to consult lists of codes in various ISO standards to find the right subtags, but now you only need to look in the IANA Language Subtag Registry.”

“In the past, when dealing with lists of ISO codes, there were sometimes multiple codes for a given language - there could be a 2-letter code and one or two 3-letter codes. This ambiguity is resolved by the IANA Subtag Registry: only one code is listed per language. (If an ISO 2-letter code exists, that will be the code, otherwise it will be a three-letter code.) The registry maintainer also coordinates the ongoing evolution of the registry with developments in the ISO world.”

<sup>2</sup> Additional sources include [Glottolog](#) and [Austlang](#).

## Proposed Changes to Controlled Lists

The following are proposed additions to controlled lists used in the DataCite Metadata Schema.

### contributorType

The [contributorType controlled list](#) is used by the Contributor sub-property [7.a contributorType](#).

#### New contributorType: Translator

**Description:** A person, organization, or automated system responsible for converting the content of a resource from one language into another, preserving its meaning and intended message.

**Usage Notes:** This contributor type should be applied to DOI metadata for a resource which has been translated from a previously published resource.

### relationType

The relationType controlled list is used by the RelatedIdentifier sub-property [12.b relationType](#) and the RelatedItem sub-property [20.b relationType](#).

#### New relationType: IsTranslationOf

**Definition:** indicates A is a translation of B

**Example:**

```
<relatedIdentifier relatedIdentifierType="DOI"
relationType="IsTranslationOf">10.21384/828a-cm38
</relatedIdentifier>
```

#### New relationType: HasTranslation

**Definition:** indicates A has a translation B

**Example:**

```
<relatedIdentifier relatedIdentifierType="DOI"
relationType="HasTranslation">10.21384/g01j-jm06</relatedIdentifier>
```

## dateType

The [dateType controlled list](#) is used by the Date sub-property [8.a dateType](#).

### New dateType: Coverage

The DataCite Metadata Schema currently lacks the concept of “temporal coverage”. This element is necessary as a standalone concept for compatibility with Schema.org’s [temporalCoverage](#) property and Dublin Core’s [temporal](#) term.

We propose to add a new dateType “Coverage” defined as:

**Description:** The date or date range that the resource content applies to, describes, or covers.

**Usage Notes:** To indicate the temporal coverage of the resource, use Coverage.

The proposed solution is designed to minimize breaking changes to the schema and achieve greater interoperability in the short-term. We acknowledge that in some domains, temporal and spatial coverage are intertwined. In the future, we may propose a combined “Coverage” property that encompasses temporal and spatial coverage.

## relatedIdentifierType

The [relatedIdentifierType controlled list](#) is used by the RelatedIdentifier sub-property [12.a relatedIdentifierType](#) and the RelatedItem sub-property [20.1.a relatedItemIdentifierType](#).

### New relatedIdentifierType: RRID

**Full Name:** Research Resource Identifier

**Description:** A character string used to uniquely identify key inputs to an experiment including the so called “key biological resources” as defined by the National Institutes of Health, and related tools. An RRID name is divided into two parts, the authority and a local identifier, separated by an underscore.

### Example:

```
<relatedIdentifier relatedIdentifierType="RRID"
relationType="Cites">RRID:AB_90755</relatedIdentifier>
```

## New relatedIdentifierType: CSTR

**Full Name:** Common Science and Technology Resources Identifier

**Description:** CSTR is an identifier based on the Chinese National Standard GB/T 32843–2016 "Science and technology resource identification", providing a unique identification service for scientific data, papers, scientific institutions, researchers, scientific instruments, patents and other scientific and technological resources.

### Example:

```
<relatedIdentifier relatedIdentifierType="CSTR"
relationType="References">31253.11.sciencedb.13238</relatedIdentifier>
```

## resourceTypeGeneral

The [resourceTypeGeneral controlled list](#) is used by the Resource Type sub-property 10.a resourceTypeGeneral. It is also used to indicate the resource type of a related resource in the RelatedIdentifier sub-property [12.f resourceTypeGeneral](#) and the RelatedItem sub-property [20.a relatedItemType](#).

## New resourceTypeGeneral: Project

**Description:** A planned endeavor or activity, frequently collaborative, intended to achieve a particular aim using allocated resources such as budget, time, and expertise.

**Examples and Usage Notes:** This resource type represents the project and includes research projects and studies. For a project deliverable or description of a project, use the corresponding resource type for the output—e.g., for a project report or study registration, use the resourceTypeGeneral "Report" or "StudyRegistration" instead.

```
<resourceType resourceTypeGeneral="Project">Masters project</resourceType>
```

```
<resourceType resourceTypeGeneral="Project">Cohort study</resourceType>
```

**Suggested Dublin Core Mapping:** N/A