

DataCite Metadata Schema 4.5: Request for Comments

Overview	2
What we are looking for	2
Ways to provide feedback	2
Comment on this summary document	2
Comment on the full draft schema	3
Other ways to share feedback	3
What happens to your feedback	4
Proposed Changes: Version 4.5	4
Support for instruments	4
Addition of Instrument to the resourceTypeGeneral controlled list values	4
Addition of new relationType pair: IsUsedBy and Uses	5
Changes and additions to definitions in support of instruments	5
Addition of new mapping: PIDINST Schema Mapping	6
Example	6
Support for pre-registrations and registration reports	7
Addition of StudyRegistration to the resourceTypeGeneral controlled list values	7
Support for publisher identifiers	8
Addition of new subproperties for Publisher	8
Example	9
New Distribution property	9
Addition of a new Distribution property	9
Addition of a new guidance document: Using Distribution for a collection of files	11
Examples	11
Clarifications to RelatedItem property	12
Changes and additions to sub-property definitions	12
Addition of a new guidance document: Using RelatedItem for publication information and related resources	13
Updated PhysicalObject definition	13
Revised definition of PhysicalObject resourceTypeGeneral	13
Other changes and corrections	13
Future Directions: Version 5.0 and beyond	14
Contributor roles	14
Related Item	15
IGSN	16

Overview

The DataCite Metadata Working Group invites your feedback on this draft proposal for Version 4.5.

These changes are in response to requests from DataCite community members: people like you that have used the metadata schema and have imagined ways in which it might work better for their particular use cases.

We want to make sure that these changes address your use cases. Therefore, for the first time, we are sharing a draft for public comment before releasing a final version Version 4.5.

This document will be open for comment through **October 17, 2022**.

What we are looking for

We are interested in your comments, questions, and suggestions about the entire metadata schema—but we are particularly focused on the Version 4.5 proposed changes.

When evaluating the Version 4.5 proposed changes, we invite you to consider these 3 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?

When evaluating the rest of the full schema draft, we are most interested in comments that address:

- Places where the schema documentation/examples are confusing, inconsistent, or otherwise ambiguous.
- Any errors or typos. This is our first time presenting the schema documentation in this format, and we may have introduced errors when transferring existing content. Let us know!

Ways to provide feedback

There are a few different ways to share feedback with us.

Comment on this summary document

This document includes the **major changes only**—not the full schema draft (it's over 100 pages now!). If you are reading this on Google Docs, you can leave comments and suggestions on:

- All major proposed changes for Version 4.5, organized by use case.
 - [Proposed Changes: Version 4.5](#) summarizes the changes under consideration for this minor version. All changes proposed for Version 4.5 are backwards compatible.
 - Each proposed change includes links to the appropriate sections of the full draft schema—so you can see the changes in context.

- Some preliminary thoughts on Version 5.0.
 - We are also starting to think about this next major release. With a major version, there is the opportunity to introduce breaking changes where required to support use cases. After careful consideration, we present two rough ideas for Version 5.0 for your input in [Future Directions: Version 5.0 and beyond](#).

If you are reading this in PDF, you can either access the Google Document here:

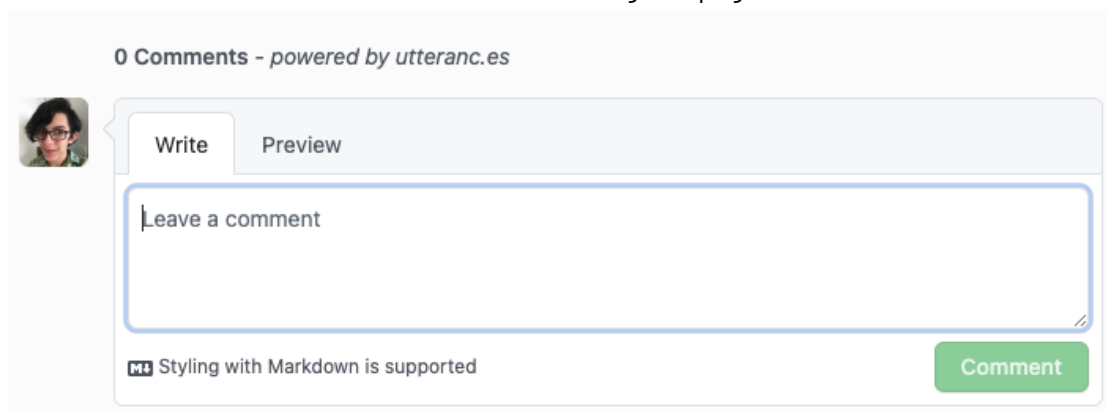
https://docs.google.com/document/d/1UyQQwtjnu-4_4zXE4TFZ74-mjLZI3NkEf8RrF0WeOdI/edit?usp=sharing or comment on the full draft schema.

Comment on the full draft schema

If you would like to comment on the **full draft schema**, it is available online here:

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

To leave feedback, scroll down to the bottom of a given page and use the comment box:



You will need a GitHub account to use this feature. How it works:

- If you are the first person to comment on a given page, your comment will open a new GitHub issue for that page.
- If there are existing comments, they will be visible on the page and you can add to the discussion.

You can use the full draft schema to comment on anything, including the Version 4.5 proposed changes.

Other ways to share feedback

Our goal is to make this feedback process as accessible and transparent as possible. If the above options do not work for you, and you would like to share your feedback publicly, please send your feedback to support@datacite.org and indicate that you would like us to post it on your behalf. You can also email us with any questions about the process.

What happens to your feedback

Your feedback on the Version 4.5 proposed changes will help us understand if the changes work for you and to make adjustments if they do not. The Metadata WG will consider the feedback and make a decision as to how to incorporate the feedback in Version 4.5, consider for future releases, or adapt feedback based on community consensus.

We will also strive to address any points of confusion, errors, and typos identified in your review of the full schema draft. Depending on the scope, some of these may be incorporated into 4.5, while others may be postponed to 5.0.

If you have shared new ideas for how we can improve the schema, we won't address these right away, but we will take them under consideration as we start planning Version 5.0. At that time, we may reach out to you with a request for more information. For this reason, we kindly request that you are logged in when commenting on this document.

Please note that while we will actively monitor comments on this document and the full schema draft, we may not respond directly to every comment.

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

Proposed Changes: Version 4.5

Note: In the case of a discrepancy between this document summarizing the proposed changes and the [full schema draft](#), the full schema draft takes precedence.

Support for instruments

Addition of [Instrument](#) to the [resourceTypeGeneral](#) controlled list values

Instrument:

Description: A device, tool or apparatus used for scientific purposes, to obtain, measure and analyze data from subjects around the research topic.

Examples and Usage Notes:

Example:

```
<resourceType  
resourceTypeGeneral="Instrument">Reflectometer</resourceType>
```

Suggested Dublin Core Mapping: N/A

Addition of new [relationType](#) pair: [IsUsedBy](#) and [Uses](#)

IsUsedBy:

Definition: Indicates A is used by B

Example and Usage Notes::

```
<relatedIdentifier relatedIdentifierType="DOI"
relationType="IsUsedBy">10.5072/dataset</relatedIdentifier>
```

May be used to indicate the relationship between an instrument and where it has been used (as in, instrument A is IsUsedBy research output B).

Uses:

Definition: Indicates A uses B

Example and Usage Notes::

```
<relatedIdentifier
relatedIdentifierType="DOI"relationType="Uses">10.5072/instrument</re
latedIdentifier>
```

May be used to indicate the relationship between an instrument and where it has been used (as in, research output A uses instrument B).

Changes and additions to definitions in support of instruments

Property	Change (changes in bold)
Title	Definition: A name or title by which a resource is known. May be the title of a dataset or the name of a piece of software or an instrument .
Creator	Definition: The main researchers involved in producing the data, or the authors of the publication, in priority order. For instruments this is the manufacturer or developer of the instrument . To supply multiple creators, repeat this property.
Contributor	<p>Definition: The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. To supply multiple contributors, repeat this property.</p> <p>For software, if there is an alternate entity that "holds, archives, publishes, prints, distributes, releases, issues, or produces" the code, use the contributorType "hostingInstitution" for the code repository.</p>

	For instruments, if there is an institution responsible for the management of the instrument (for example, the legal owner, the operator, or an institute providing access to the instrument), use the contributorType "hostingInstitution" for the owner of the instrument.
AlternateIdentifier	Definition: An identifier other than the primary Identifier applied to the resource being registered. This may be any alphanumeric string which is unique within its domain of issue. May be used for local identifiers, a serial number of an instrument or an inventory number . The AlternateIdentifier should be an additional identifier for the same instance of the resource (i.e., same location, same file).
Description	Definition: All additional information that does not fit in any of the other categories. May be used for technical information or detailed information associated with a scientific instrument .
descriptionType: TechnicalInfo	<p>Definition: Detailed information that may be associated with design, implementation, operation, use, and/or maintenance of a process, system, or instrument.</p> <p>Usage Notes: For software description, this may include the contents of a "readme.txt" and necessary environmental information (hardware, operational software, applications/programs with version information, a human-readable synopsis of software purpose) that cannot be described using other properties (e.g., programming language). For other uses, this can include specific and detailed information as necessary and appropriate. The information entered will be unstructured and not parsed, so it may be useful to format the information using line breaks ("
").</p>

Addition of new mapping: PIDINST Schema Mapping

→ Please visit [PIDINST Schema Mapping](#) to comment on this documentation change.

Example

```
<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
https://schema.datacite.org/meta/kernel-4.4/metadata.xsd">
  <identifier identifierType="DOI">10.5072/example-instrument</identifier>
  <creators>
    <creator>
      <creatorName nameType="Organizational">DECTRIS</creatorName>
      <nameIdentifier schemeURI="https://www.wikidata.org/"
nameIdentifierScheme="Wikidata">Q107529885</nameIdentifier>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en-US">FPilatus detector at MX station 14.1</title>
  </titles>
  <publisher xml:lang="en">Helmholtz-Zentrum Berlin für Materialien und Energie</publisher>
  <publicationYear>2021</publicationYear>
  <contributors>
    <contributor contributorType="HostingInstitution">
```

```
<contributorName nameType="Organizational">Helmholtz-Zentrum Berlin für Materialien und
Energie</contributorName>
<nameIdentifier schemeURI="https://ror.org/"
nameIdentifierScheme="ROR">https://ror.org/02aj13c28</nameIdentifier>
</contributor>
</contributors>
<resourceType resourceTypeGeneral="Instrument">Raster image pixel detector</resourceType>
<alternateIdentifiers>
<alternateIdentifier alternateIdentifierType="SerialNumber">1234567</alternateIdentifier>
</alternateIdentifiers>
<relatedIdentifiers>
<relatedIdentifier relatedIdentifierType="Handle" relationType="IsPartOf"
resourceTypeGeneral="Instrument">1234.1675</relatedIdentifier>
<relatedIdentifier relatedIdentifierType="URL" relationType="References"
resourceTypeGeneral="Text">https://www.dectris.com/products/pilatus3/pilatus3-s-for-synchrotron/det
ails/pilatus3-s-6m</relatedIdentifier>
</relatedIdentifiers>
<descriptions>
<description xml:lang="en-US" descriptionType="Abstract">The Pilatus 6M pixel-detector at the
MX station 14.1</description>
<description xml:lang="en-US" descriptionType="TechnicalInfo">Model Name: PILATUS3 S 6M.
Instrument type: Raster image pixel detector. Measured variables: X-ray.</description>
</descriptions>
</resource>
```

Support for pre-registrations and registration reports

Addition of [StudyRegistration](#) to the [resourceTypeGeneral](#) controlled list values

StudyRegistration:

Description: A detailed, time-stamped description of a 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.

Examples and Usage Notes: Includes pre-registrations, registered reports, and clinical trials. Study registrations are sometimes peer-reviewed and may include the hypothesis, expected results, study design, and/or analysis plan.

Example:

```
<resourceType
resourceTypeGeneral="StudyRegistration">Pre-registration</resourceType>
```

Suggested Dublin Core Mapping: Text

Support for publisher identifiers

Addition of new subproperties for [Publisher](#)

[4.a publisherIdentifier](#)

Occurrences: 0-1

Definition: Uniquely identifies the publisher, according to various schemes.

Allowed values, examples, other constraints:

Examples:

- <https://doi.org/10.17616/R3989R>
- <https://ror.org/04z8jg394>
- <https://viaf.org/viaf/151411898/>
- <https://www.wikidata.org/wiki/Q7842>

[4.b publisherIdentifierScheme](#)

Occurrences: 1

Definition: The name of the publisher identifier scheme

Allowed values, examples, other constraints:

If publisherIdentifier is used, publisherIdentifierScheme is mandatory.

Examples:

- re3data
- ROR
- VIAF
- Wikidata
- Crossref Funder ID
- ISNI
- OpenDOAR
- FAIRsharing
- Ringgold
- ISSN

[4.c schemeURI](#)

Occurrences: 0-1

Definition: The URI of the publisher identifier scheme

Allowed values, examples, other constraints:

Examples:

- <https://www.re3data.org/>
- <https://ror.org/>
- <https://viaf.org/>
- <https://www.wikidata.org>

Example

```
<publisher xml:lang="en" publisherIdentifier="http://doi.org/10.17616/R3J014"
publisherIdentifierScheme="re3data" schemeURI="https://www.re3data.org">Global Biodiversity
Information Facility</publisher>
```

New Distribution property

Addition of a new Distribution property

[21. Distribution](#)

Occurrences: 0-n

Definition: Represents an accessible form of a resource such as downloadable files.

Allowed values, examples, other constraints:

The use of this property indicates directly downloadable distributions. Every distribution should represent the same resource in its entirety. It should NOT be used to describe collections.

Collections of files should be either using an archive format or a BagIt folder structure. See [Using Distribution for a collection of files](#) for recommendations.

[21.a mediaType](#)

Occurrences: 1

Definition: Media type expressed using a MIME format.

Allowed values, examples, other constraints:

If Distribution is used, mediaType is mandatory.

Only MIME formats are allowed: see [IANA site](#) and [MDN reference](#).

Examples:

- application/zip
- audio/mpeg

[21.1 contentURL](#)

Occurrences: 1

Definition: The URL leading to content provided by a repository using a valid protocol.

Allowed values, examples, other constraints:

If Distribution is used, contentURL is mandatory.

URLs should use schemes that are registered with IANA (e.g., https, ftp):

<https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml>

Examples:

- <https://example.org/data.csv>
- <ftp://example.org/data.txt>
- <https://example.org/bagit.zip>

- <https://example.org/files.gzip>

See [Using Distribution for a collection of files](#) for recommendations on archive file formats.

[21.1.a lastUpdated](#)

Occurrences: 0-1

Definition: Date when the content URL value was last updated.

Allowed values, examples, other constraints:

YYYY, YYYY-MM-DD, YYYYMM-DDThh:mm:ssTZD or any other format or level of granularity described in W3CDTF.

[21.1.b byteSize](#)

Occurrences: 0-1

Definition: The size of a distribution in bytes. The value is related to the object in 21.1 contentURL.

Allowed values, examples, other constraints:

The size in bytes can be approximated (as a decimal) when the precise size is not known.

Examples: - 1048576 for 1 Megabyte

[21.2 checksum](#)

Occurrences: 0-1

Definition: A value that allows the integrity of a file to be verified. The value is related to the object in 21.1 contentURL.

Allowed values, examples, other constraints:

This attribute allows the results of a variety of checksum and cryptographic message digest algorithms to be represented.

[21.2.a algorithm](#)

Occurrences: 1

Definition: Identifies the algorithm used to produce the checksum.

Allowed values, examples, other constraints:

If checksum is used, algorithm is mandatory.

Recommended values should follow Version 2.3 of SPDX: <https://spdx.org/rdf/terms/#d4e1968>

Examples:

- MD2
- MD4
- MD5
- MD6
- SHA-1
- SHA-224
- SHA-256
- SHA-384

- SHA-512

[21.3 accessRights](#)

Occurrences: 0-1

Definition: A rights statement that describes how the distribution is accessed.

Allowed values, examples, other constraints:

Recommended values should follow the COAR vocabulary to declare the access status of a resource:

https://vocabularies.coar-repositories.org/access_rights/

To provide copyright or licensing information, use the [Rights](#) property.

Examples:

- embargoed access
- metadata only access
- open access

[21.3.a accessRightsUri](#)

Occurrences: 0-1

Definition: The URI used to define the access rights.

Allowed values, examples, other constraints:

Recommended values should follow the COAR vocabulary for to declare the access status of a resource:

https://vocabularies.coar-repositories.org/access_rights/

Examples:

- http://purl.org/coar/access_right/c_abf2 for "open access"
- http://purl.org/coar/access_right/c_14cb for "metadata only"

Addition of a new guidance document: Using Distribution for a collection of files

→ Please visit [Using Distribution for a collection of files](#) to comment on this documentation change.

Examples

Multiple files for the digital object with BagIt:

```
<distributions>
  <distribution mediaType="application/gzip">
    <contentURL lastUpdated="2022-05-05"
byteSize="1236546456">https://example.org/bagit.gzip</contentURL>
    <checksum algorithm="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
    <accessRights xml:lang="en" accessRightsUri="http://purl.org/coar/access_right/c_abf2">open
access</accessRights>
  </distribution>
</distributions>
```

Multiple formats of the same digital object:

```
<distributions>
  <distribution mediaType="image/png">
    <contentURL lastUpdated="2022-05-05"
byteSize="1236546456">http://example.org/data.png</contentURL>
    <checksum algorithm="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
    <accessRights xml:lang="en"
accessRightsUri="http://purl.org/coar/access_right/c_flcf">embargoed access</accessRights>
  </distribution>
  <distribution mediaType="text/csv">
    <contentURL>http://example.org/data.csv</contentURL>
  </distribution>
</distributions>
```

Clarifications to RelatedItem property

Changes and additions to sub-property definitions

Property	Change
20.1 relatedItemIdentifier	Add note: If relatedItemIdentifier is provided, an identical 12. RelatedIdentifier is strongly recommended for indexing.
20.5 volume	Add note: Use only with relationType IsPublishedIn .
20.6 issue	
20.7 number	
20.7.a numberType	
20.8 firstPage	
20.9 lastPage	
20.11 edition	
20.8 firstPage	Change definitions to specify that the pages refer to the resource within the related item (for which the DOI is being registered), not the entire related item. 20.8 firstPage: First page of the resource within the related item, e.g., of the chapter, article, or conference paper in proceedings 20.9 lastPage: Last page of the resource within the related item, e.g., of the chapter, article, or
20.9 lastPage	

	conference paper in proceedings
--	---------------------------------

The definition of descriptionType [SeriesInformation](#) in [17.a descriptionType](#) and [Appendix 1: Controlled List Definitions - descriptionType](#) was also updated to clarify this descriptionType is superseded by RelatedItem with the relationType [IsPublishedIn](#) selected.

Addition of a new guidance document: Using RelatedItem for publication information and related resources

→ Please visit [Using RelatedItem for publication information and related resources](#) to comment on this documentation change.

Updated PhysicalObject definition

Revised definition of [PhysicalObject](#) resourceTypeGeneral

Revised:

Definition: A physical object or substance.

Usage notes: Artifacts, specimens, material samples, and features-of-interest of any size. Note that digital representations of physical objects should use one of the other resourceTypeGeneral values.

Previous:

Definition: An inanimate, three-dimensional object or substance.

Usage notes: Artifacts, specimens

Other changes and corrections

- Correction of the cardinality for properties [2.5.a affiliationIdentifier](#) and [7.5.a affiliationIdentifier](#).
- Correction of the capitalization of properties [2.5.c schemeURI](#), [7.5.c schemeURI](#), and [19.2.b schemeURI](#).
- Correction of the cardinality of [19.2.a funderIdentifierType](#) and addition of a note to indicate when it is mandatory.
- Addition of a note to [3.a titleType](#) (sub-property of [3. Title](#)) to match the corresponding note in [20.3.a titleType](#) (subproperty of [20.3 Title](#) in [20. RelatedItem](#)).
- Updated examples for nameIdentifier (properties [2.4](#) and [7.4](#)) and its attributes.
- Updated examples for affiliationIdentifier (properties [2.5](#) and [7.5](#)) and its attributes.
- Updated the full name of [relatedIdentifierType IGSN](#) from "International Geo Sample Number" to "International Generic Sample Number" with an updated description.
- Other minor corrections to definitions and examples.

Future Directions: Version 5.0 and beyond

Contributor roles

A revision to better support contributor roles is under consideration for the next major version (5.0).

For version 4.5, the Metadata Working Group considered how to allow for interoperability between the current contributorType list and the [CRediT Contributor Roles Taxonomy](#). However, due to issues around any 1-1 mapping of each lists' values and the existing limitations of the contributorType attribute, we concluded that this was not possible without introducing a breaking change.

ContributorType is an XML attribute and therefore cannot be made repeatable, which is an option with the CRediT and other vocabulary values. Since there is an existing DataCite controlled list for contributor types any decision to introduce external vocabularies requires careful consideration of what happens to the existing contributorType controlled list values.

As we look ahead to version 5.0, the Metadata Working Group is considering how to allow for greater interoperability with external vocabularies across the entire metadata schema. This is already in place for some properties: for example, the Subject property which has attributes for subjectScheme, schemeUri, valueUri and classificationCode. This format could be reproducible across relevant vocabularies in the DataCite schema.

The Metadata Working Group is therefore considering the following breaking changes for version 5.0:

- Add a repeatable **role** sub-element to the **Contributor** and **Creator** properties
 - Attributes: **roleScheme**, **schemeURI**, **valueURI**

Example:

```
<contributor>
  <contributorName nameType="Personal">Garcia, Sofia</contributorName>
    <givenName>Sofia</givenName>
    <familyName>Garcia</familyName>
    <nameIdentifier schemeURI="https://orcid.org/"
      nameIdentifierScheme="ORCID">0000-0001-5727-2427</nameIdentifier>
    <affiliation affiliationIdentifier="https://ror.org/03efmqc40" affiliationIdentifierScheme="ROR"
      schemeURI="https://ror.org">Arizona State University</affiliation>
    <roles>
      <role roleScheme="CRediT" schemeURI="https://credit.niso.org/contributor-roles/"
        valueURI="https://credit.niso.org/contributor-roles/conceptualization/">Conceptualization</role>
      <role roleScheme="CRediT" schemeURI="https://credit.niso.org/contributor-roles/"
        valueURI="https://credit.niso.org/contributor-roles/methodology/">Methodology</role>
    </roles>
  </contributor>
```

</contributor>

With this proposal, we would also revise and publish the existing DataCite contributorTypes as one such vocabulary that could be referenced from the **role** sub-element, either independently or in addition to other roles. For example:

```
<roles>
  <role roleScheme="CRediT" schemeURI="https://credit.niso.org/contributor-roles/"
    valueURI="https://credit.niso.org/contributor-roles/methodology/">Methodology</role>
  <role roleScheme="DataCite contributorTypes" schemeURI="[ to be determined]" valueURI="[ to be
    determined]">Researcher</role>
</roles>
```

The Metadata Working Group welcomes your feedback on this approach and is interested in your ideas for how we can better support contributor roles in version 5.0.

Related Item

A revision of the RelatedItem property is under consideration for the next major version (5.0).

The RelatedItem property was originally developed to satisfy two distinct use cases:

The first use case is providing publication information for journal articles, book chapters, and other resources that are published within another item. This information about the related item (the container) is needed to formulate a complete citation of the primary resource being described. For example, a book title is necessary to cite a book chapter, and a journal title and volume/issue number are necessary to cite a journal article.

The second use case is providing information about related resources.

- When a related resource does not have an identifier, the [20. RelatedItem](#) property should be used to provide information about the related resource.
- When a related resource has an identifier, the [12. RelatedIdentifier](#) property should always be used. In addition, the [20. RelatedItem](#) property may optionally be used to provide information about the related resource.

For version 4.5, the Metadata Working Group has developed additional guidance for how to apply the RelatedItem property to these use cases: [Using RelatedItem for publication information and related resources](#)

The Metadata Working Group welcomes your ideas on how to better support these use cases in schema version 5.0.

IGSN

A revision to better support IGSN IDs is under consideration for the next major version (5.0).

In 2021, DataCite and the International Generic Sample Number (IGSN) e.V. announced a partnership to support the global adoption, implementation, and use of physical sample identifiers. Under the partnership, DataCite is providing the IGSN ID registration services and supporting technology to enable the ongoing sustainability of the IGSN ID PID infrastructure. Consequently, IGSN IDs registered with DataCite adopt the DataCite Metadata Schema.

Version 4.5 contains a revision to the PhysicalObject definition to better meet the needs of material samples communities; specifically, those that deal with living samples. The Metadata WG updated this definition at the request of the IGSN–DataCite Crosswalk and Metadata Management WG, a partnership group of the IGSN e.V. and DataCite.

For version 5.0, the Metadata WG will consider ways to more effectively support IGSN IDs and metadata for physical samples, and we welcome feedback and suggestions to this end. One important consideration that may be included is a way to differentiate IGSN IDs from other DOIs at the metadata level. There may also be a need to better record samples that have been destroyed or discarded during the analysis process.