



DataCite - International Data Citation

DataCite Metadata Schema for the Publication and Citation of Research Data

Version 3.0

July 2013

doi:10.5438/0008

Members of the Metadata Working Group

Joan Starr, California Digital Library (chair of working group)

Jan Ashton, British Library

Amy Barton, Purdue University Library

Jannean Elliott, DOE/OSTI

Marie-Christine Jacquemot-Perbal, Inist-CNRS

Merja Karjalainen, SND

Lynne McAvoy, CISTI

Elizabeth Newbold, British Library

Sebastian Peters, TIB

Madeleine de Smaele, TU Delft

Natalija Schleinstein, GESIS

Wolfgang Zenk-Möltgen, GESIS

Frauke Ziedorn, TIB (Metadata Supervisor)

Table of Contents

1	Introduction.....	3
1.1	The DataCite Consortium	3
1.2	DataCite Community Participation	3
1.3	The Metadata Schema	3
1.4	Version 3 Update	4
2	DataCite Metadata Properties	5
2.1	Overview	5
2.2	Citation	7
2.3	DataCite Properties	8
3	XML Example	18
4	XML Schema	18
5	Other DataCite Services	19
	Appendices	20
	Appendix 1: Controlled List Definitions.....	20
	Appendix 2: Earlier Version Update Notes.....	32

1 Introduction

1.1 The DataCite Consortium

Scholarly research is producing ever increasing amounts of digital research data, and it depends on data to verify research findings, create new research, and share findings. In this context, what has been missing until recently, is a *persistent* approach to access, identification, sharing, and re-use of datasets. To address this need, the DataCite¹ international consortium was founded in late 2009 with these three fundamental goals:

- establish easier access to scientific research data on the Internet,
- increase acceptance of research data as legitimate, citable contributions to the scientific record, and
- support data archiving that will permit results to be verified and re-purposed for future study.

Since its founding in 2009, DataCite has grown and now spans the globe from Europe and North America to Asia and Australia. The aim of DataCite is to provide domain agnostic services to benefit scholars in a wide range of disciplines.

Key to the DataCite service is the concept of a long term or *persistent* identifier. A persistent identifier is an association between a character string and a resource. Resources can be files, parts of files, persons, organizations, abstractions, etc. DataCite uses *Digital Object Identifiers* (DOIs)² at the present time and is investigating the use of other identifier schemes in the future. For this reason, the Metadata Schema has been designed with flexibility and extensibility in mind.

1.2 DataCite Community Participation

The Metadata Working Group would like to acknowledge the contributions to our work of many colleagues in our institutions who provided assistance of all kinds. Their help has been greatly appreciated. In addition, we are indebted to numerous individuals and organizations in the broader scholarly community who have taken an interest in this work. Because data citation and data management are evolving areas of concern, we look forward to continued interest. With this in mind, we have laid the groundwork for increased interaction within the working group, and between the group and the DataCite Community, and beyond.

1.3 The Metadata Schema

The DataCite Metadata Schema is a list of core metadata properties chosen for the accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. The resource that is being identified can be of any kind, but it is typically a dataset. We use the term 'dataset' in its broadest sense. We mean it to include not only numerical data, but any other research data outputs. The metadata schema properties are presented and described in detail in Section 2.

In this release of the metadata schema, there are some larger changes. The documentation now indicates which optional properties can be used to achieve greater exposure for the resource's

¹ <http://www.datacite.org>

² DOIs are administered by the International DOI Foundation, <http://www.doi.org/>

metadata (see Section 2.1). Another notable change is that there is a new option available: the ability to link external, local, and /or domain-specific metadata to the DataCite record. We will observe with interest how DataCite clients make use of this new facility. For a specified list of changes, see Section 1.4.

Please note that DataCite reserves the right to share metadata with information indexes and other entities.³

Lastly, in order to support openness and future extensibility of the schema, a collaboration between DataCite and the Dublin Core Metadata Initiative (DCMI) Science and Metadata Community (SAM)⁴ is producing a version of the schema in a Dublin Core Application Profile format. The Application Profile version will be released following this Version 3.0 release and will be maintained with all future versions of the Schema.

1.4 Version 3 Update

Version 3.0 of the DataCite Metadata Schema includes these changes⁵:

- Correction of a problem with our way of depicting dates by
 - implementing RKMS-ISO8601⁶ standard for depicting date ranges, so that a range is indicated as follows: 2004-03-02/2005-06-02
 - deleting `startDate` and `endDate` date types, and derogating these from earlier versions
- Addition of a new `GeoLocation` property, with the sub-properties `geoLocationPoint`, `geoLocationBox`, `geoLocationPlace` supporting a simple depiction of geospatial information, as well as a free text description.
- Addition of new values to controlled lists:
 - `contributorType`: `ResearchGroup` and `Other`
 - `dateType`: `Collected`
 - `resourceTypeGeneral`: `Audiovisual`, `Workflow`, and `Other` and derogation of `Film`
 - `relatedIdentifierType`: `PMID`
 - `relationType`: `IsIdenticalTo` (indicates that A is identical to B, for use when there is a need to register two separate instances of the same resource)
 - `relationType`: `HasMetadata`, (indicates resource A has additional metadata B and indicates), `IsMetadataFor` (indicates additional metadata A for resource B)
 - `descriptionType`: `Methods`
- Deletion of the derogated `resourceType`: `film`
- new sub-properties for `relationType`: `relatedMetadataScheme` and `schemeURI`, to be used only for the new `relationType` pair of `HasMetadata`, `IsMetadataFor`
- Addition of `schemeURI` sub-property to the `nameIdentifierScheme` associated with `CreatorName`, `ContributorName` and `Subject`
- Addition of `theschemeURI` sub-property to `Rights`

³ This is addressed in the DataCite Business Principles, available here:

http://datacite.org/sites/default/files/Business_Models_Principles_v1.0.pdf

⁴ For more information on DCMI SAM, see http://wiki.dublincore.org/index.php/DCMI_Science_And_Metadata.

⁵ Two additional schema code level changes are the allowance of keeping optional wrapper elements empty and the allowance of arbitrary ordering of elements (by removal of `<xs:sequence>`).

⁶ The standard is documented here: <http://www.ukoln.ac.uk/metadata/dcml/collection-RKMS-ISO8601/>

- Implementation of the `xml:lang` attribute⁷ that can be used on the properties `Title`, `Subject` and `Description`.
- Removal of two system-generated administrative metadata fields: `LastMetadataUpdate` and `MetadataVersionNumber` because both values are tracked in another way now.

Version 3.0 of the DataCite Metadata Schema documentation includes these changes:

- Updates to the introductory information
- Provision of greater detail, explanatory material and definitions for controlled lists
- Indication of recommended metadata, in addition to mandatory and optional
- Addition of more and more varied XML examples on the Metadata Schema website
- Removal from documentation of information about administrative metadata (which cannot be edited by contributors).

2 DataCite Metadata Properties

2.1 Overview

The properties of the DataCite Metadata Schema are presented in this section. More detailed descriptions of the properties, and their related sub-properties, are provided in Sections 2.3 and 2.4.

There are three different levels of obligation for the metadata properties:

- **Mandatory (M)** properties *must* be provided,
- **Recommended (R)** properties are optional, but strongly recommended and
- **Optional (O)** properties are optional and provide richer description.

Those clients who wish to enhance the prospects that their metadata will be found, cited and linked to original research are strongly encouraged to submit the Recommended as well as Mandatory set of properties. Together, the Mandatory and Recommended set of properties and their sub-properties are especially valuable to information seekers and added-service providers, such as indexers. The Metadata Working Group members strongly urge the inclusion of metadata identified as Recommended for the purpose of achieving greater exposure for the resource's metadata record, and therefore, the underlying research itself.

The properties listed in Table 1 have the obligation level Mandatory, and *must be* supplied when submitting DataCite metadata. The properties listed in Table 2 have one of the obligation levels Recommended or Optional, and *may be* supplied when submitting DataCite metadata.

In Tables 1 and 2, the combined Mandatory and Recommended “super set” of properties and sub-properties that enhance the prospect that the resource's metadata will be found, cited and linked is indicated by shading, as shown in the example below.

Example of shading

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
1	Identifier	1	The Identifier is a unique string that identifies a resource.	DOI (Digital Object Identifier) registered by a DataCite

⁷ Allowed values IETF BCP 47, ISO 639-1 language codes, e.g. en, de, fr

			member. Format should be "10.1234/foo"
--	--	--	--

Of the Recommended set of properties, the two most important to use are the properties `ResourceType` and `Description`, together with their Recommended sub-properties `resourceTypeGeneral="controlled list value"` (see Appendix 1) and `descriptionType="Abstract"` (see Section 2.3). Appendix 1 includes detailed descriptions of controlled list values, using the same shading to indicate those values that are especially important for information seekers and added-service providers.

It cannot be emphasized enough how valuable `resourceTypeGeneral` and `Description` are to other scholars in finding the resource and then determining whether or not the resource, once found, is worth investigating further.

Table 1: DataCite Mandatory Properties

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
1	Identifier (with type sub-property)	M
2	Creator (with name identifier sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M

Table 2: DataCite Recommended and Optional Properties

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
6	Subject (with scheme sub-property)	R
7	Contributor (with type and name identifier sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
10	ResourceType (with general type description sub-property)	R
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point and box sub-properties)	R

2.2 Citation

Because many users of this schema are members of a variety of academic disciplines, DataCite remains discipline-agnostic concerning matters pertaining to academic style sheet requirements. Therefore, DataCite encourages rather than requires a particular citation format⁸. In keeping with this approach, the following is the *preferred* format for rendering a DataCite citation for human readers using the first five properties of the schema:

Creator (PublicationYear): Title. Publisher. Identifier

It may also be desirable to include information from two optional properties, Version and Resource Type (as appropriate). If so, the preferred form is as follows:

Creator (PublicationYear): Title. Version. Publisher. Resource Type. Identifier

For citation purposes, DataCite prefers that DOI names are displayed as linkable, permanent URLs. The Identifier may appear in its original format. If the original format is chosen, be sure to include the characters "doi:" pre-pended to the Identifier as in "doi:10.1234/abc."

For resources that do not have a standard publication year value, DataCite suggests that PublicationYear should include the date that would be preferred for use in a citation.

Here are several examples:

- Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V. 2.1. Geological Institute, University of Tokyo.
<http://dx.doi.org/10.1594/PANGAEA.726855>
- Geofon operator (2009): GEFON event gfz2009kciu (NW Balkan Region). GeoForschungsZentrum Potsdam (GFZ). <http://dx.doi.org/10.1594/GFZ.GEOFON.gfz2009kciu>
- Denhard, Michael (2009): dphase_mpeps: MicroPEPS LAF-Ensemble run by DWD for the MAP D-PHASE project. World Data Center for Climate.
http://dx.doi.org/10.1594/WDCC/dphase_mpeps

A special note regarding citation of dynamic datasets:

For datasets that are continuously and rapidly updated, there are special challenges both in citation and preservation. For citation, three approaches are possible:

- a) Cite a specific time slice (the set of updates to the dataset made during a particular period of time);
- b) Cite a specific snap shot (a copy of the entire dataset made at a specific time);
- c) Cite the continuously updated dataset, but add an Access Date and Time to the citation.

Note that a "time slice" and "snap shot" are versions of the dataset and require unique identifiers. The third option is controversial, because it necessarily means that following the citation does not result in observation of the resource as cited.

⁸ In collaboration with CrossRef, DataCite has created a DOI Citation Formatter Service available at <http://crosscite.org/citeproc/>. The user can choose from more than 500 different citation formats in 45 different languages.

2.3 DataCite Properties

Table 3 provides a detailed description of the mandatory properties, together with their sub-properties, which *must* be supplied with any initial metadata submission to the managing agent for DataCite. In Table 4, the Recommended and Optional properties are described in detail. For an example of how to make a submission in XML format, please see the XML Examples provided on the DataCite Metadata Schema Repository⁹ website.

Throughout this document, a naming convention has been used for all properties and sub-properties as follows: properties begin with a capital letter, whereas sub-properties begin with a lower case letter. If the name is a compound of more than one word, subsequent words begin with capital letters.

As with Tables 1 and 2, Tables 3 and 4 use shading to identify the combined Mandatory and Recommended “super set” of properties and sub-properties that enhance the prospect that the resource’s metadata will be found, cited and linked.

The third column, Occurrence (Occ), indicates cardinality/quantity constraints for the properties as follows:

- 0-n = optional and repeatable
- 0-1 = optional, but not repeatable
- 1-n = required and repeatable
- 1 = required, but not repeatable

NOTE:

XML provides an `xml:lang` attribute¹⁰ that can be used on the properties `Title`, `Subject` and `Description`. This provides a way to describe the language used for the content of the specified properties. The schema provides a `Language` property to be used to describe the language of the resource.

⁹ <http://schema.datacite.org/>

¹⁰ Allowed values IETF BCP 47, ISO 639-1 language codes, e.g. en, de, fr

Table 3: Expanded DataCite Mandatory Properties

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
1	Identifier	1	The Identifier is a unique string that identifies a resource.	DOI (Digital Object Identifier) registered by a DataCite member. Format should be “10.1234/foo”
1.1	identifierType	1	The type of the Identifier.	<i>Controlled List Value:</i> DOI
2	Creator	1-n	The main researchers involved in producing the data, or the authors of the publication, in priority order.	May be a corporate/institutional or personal name.
2.1	creatorName	1	The name of the creator.	Examples: Smith, John; Miller, Elizabeth The personal name format should be: family, given. Non-roman names may be transliterated according to the ALA-LC schemes ¹¹ .
2.2	nameIdentifier	0-1	Uniquely identifies an individual or legal entity, according to various schemes.	The format is dependent upon scheme.
2.2.1	nameIdentifierScheme	1	The name of the name identifier scheme.	If nameIdentifier is used, nameIdentifierScheme is mandatory. Examples: ORCID ¹² , ISNI ¹³ ,
2.2.2	schemeURI	0-1	The URI of the name identifier scheme.	Examples : http://www.isni.org or http://www.orcid.org
3	Title	1-n	A name or title by which a resource is known.	Free text.
3.1	titleType	0-1	The type of Title.	<i>Controlled List Values:</i> AlternativeTitle Subtitle TranslatedTitle

¹¹ <http://www.loc.gov/catdir/cpsd/roman.html>

¹² <http://www.orcid.org/>

¹³ <http://www.isni.org/>

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
4	Publisher	1	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role.	Examples: World Data Center for Climate (WDCC); GeoForschungsZentrum Potsdam (GFZ); Geological Institute, University of Tokyo
5	PublicationYear	1	The year when the data was or will be made publicly available.	<p>YYYY</p> <p>***</p> <p>If an embargo period has been in effect, use the date when the embargo period ends.</p> <p>In the case of datasets, "publish" is understood to mean making the data available on a specific date to the community of researchers.</p> <p>If there is no standard publication year value, use the date that would be preferred from a citation perspective.</p>

Table 4: Expanded DataCite Recommended and Optional Properties

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
6	Subject	0-n	Subject, keyword, classification code, or key phrase describing the resource.	Free text.
6.1	subjectScheme	0-1	The name of the subject scheme or classification code or authority if one is used.	Free text.
6.2	schemeURI	0-1	The URI of the subject identifier scheme.	Examples: http://id.loc.gov/authorities/subjects/ ; http://dewey.info/
7	Contributor	0-n	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.	

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
7.1	contributorType	1	The type of contributor of the resource.	<p>If Contributor is used, then contributorType is mandatory.</p> <p><i>Controlled List Values:</i></p> <p>ContactPerson DataCollector DataManager Distributor Editor Funder HostingInstitution Producer ProjectLeader ProjectManager ProjectMember RegistrationAgency RegistrationAuthority RelatedPerson Researcher ResearchGroup RightsHolder Sponsor Supervisor WorkPackageLeader Other</p> <p>See appendix for definitions.</p>
7.2	contributorName	1	The name of the contributor.	<p>If Contributor is used, then contributorName is mandatory.</p> <p>Examples: Patel, Emily; Doe, John</p> <p>The personal name format may be: family, given. Non-roman names should be transliterated according to the ALA-LC schemes¹⁴.</p>
7.3	nameIdentifier	0-1	Uniquely identifies an individual or legal entity, according to various schemes.	The format is dependent upon scheme.
7.3.1	nameIdentifierScheme	1	The name of the name identifier scheme.	<p>If nameIdentifier is used, nameIdentifierScheme is mandatory.</p> <p>Examples: ORCID¹⁵, ISNI¹⁶</p>

¹⁴ <http://www.loc.gov/catdir/cpsd/roman.html>

¹⁵ <http://www.orcid.org/>

¹⁶ <http://www.isni.org/>

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
7.3.2	schemeURI	0-1	The URI of the name identifier scheme.	Examples: http://www.isni.org http://www.orcid.org
8	Date	0-n	Different dates relevant to the work.	YYYY,YYYY-MM-DD, YYYY-MM-DDThh:mm:ssTZD or any other format or level of granularity described in W3CDF. ¹⁷ Use RKMS-ISO8601 ¹⁸ standard for depicting date ranges. Example: 2004-03-02/2005-06-02
8.1	dateType	1	The type of date.	If Date is used, dateType is mandatory. <i>Controlled List Values:</i> Accepted Available Copyrighted Collected Created Issued Submitted Updated Valid See appendix for definitions.
9	Language	0-1	The primary language of the resource.	Allowed values are taken from IETF BCP 47, ISO 639-1 language codes. Examples: en, de, fr
10	ResourceType	0-1	A description of the resource.	The format is open, but the preferred format is a single term of some detail so that a pair can be formed with the sub-property. *** Example: Image/Animation, where 'Image' is resourceTypeGeneral value and 'Animation' is ResourceType value.

¹⁷ <http://www.w3.org/TR/NOTE-datetime>

¹⁸ The standard is documented here: <http://www.ukoln.ac.uk/metadata/dcmi/collection-RKMS-ISO8601/>

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
10.1	resourceTypeGeneral	1	The general type of a resource.	<p>If ResourceType is used, resourceTypeGeneral is mandatory.</p> <p><i>Controlled List Values:</i></p> <p>Audiovisual Collection Dataset Event Image InteractiveResource Model PhysicalObject Service Software Sound Text Workflow Other</p> <p>See appendix for definitions and examples.</p>
11	AlternateIdentifier	0-n	An identifier or identifiers 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. AlternateIdentifier should be used for another identifier of the same instance (same location, same file).	<p>Free text.</p> <p>***</p> <p>Example:</p> <p>A local accession number: E-GEOD-34814</p>
11.1	alternateIdentifierType	1	The type of the AlternateIdentifier.	<p>Free text.</p> <p>***</p> <p>If AlternateIdentifier is used, alternateIdentifierType is mandatory.</p>
12	RelatedIdentifier	0-n	Identifiers of related resources. These must be globally unique identifiers.	<p>Free text.</p> <p>***</p> <p>Use this property to indicate subsets of properties, as appropriate.</p>

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
12.1	relatedIdentifierType	1	The type of the RelatedIdentifier	<p>If RelatedIdentifier is used, relatedIdentifierType is mandatory.</p> <p><i>Controlled List Values:</i></p> <p>ARK DOI EAN13 EISSN Handle ISBN ISSN ISTC LISSN LSID PMID PURL UPC URL URN</p> <p>See appendix for full names and examples.</p>
12.2	relationType	1	Description of the relationship of the resource being registered (A) and the related resource (B).	<p>If RelatedIdentifier is used, relationType is mandatory.</p> <p><i>Controlled List Values:</i></p> <p>IsCitedBy Cites IsSupplementTo IsSupplementedBy IsContinuedBy Continues HasMetadata IsMetadataFor IsNewVersionOf IsPreviousVersionOf IsPartOf HasPart IsReferencedBy References IsDocumentedBy Documents IsCompiledBy Compiles IsVariantFormOf IsOriginalFormOf IsIdenticalTo</p> <p>See appendix for definitions, examples and usage notes.</p>

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
12.3	relatedMetadataScheme	0-1	The name of the scheme.	Use only with this relation pair: (HasMetadata/ IsMetadataFor)
12.4	schemeURI	0-1	The URI of the relatedMetadataScheme.	Use only with this relation pair: (HasMetadata/ IsMetadataFor)
12.5	schemeType	0-1	The type of the relatedMetadataScheme	Use only with this relation pair: (HasMetadata/ IsMetadataFor) Examples: XSD, DDT, Turtle
13	Size	0-n	Unstructured size information about the resource.	Free text. *** Examples: "15 pages", "6 MB"
14	Format	0-n	Technical format of the resource.	Free text. *** Use file extension or MIME type where possible, e.g., PDF, XML, MPG or application/pdf, text/xml, video/mpeg.
15	Version	0-1	The version number of the resource.	Suggested practice: track major_version.minor_version. Register a new identifier for a major version change. Individual stewards need to determine which are major vs. minor versions ¹⁹ . May be used in conjunction with properties 11 and 12 (AlternateIdentifier and RelatedIdentifier) to indicate various information updates. May be used in conjunction with property 17 (Description) to indicate the nature and file/record range of version.

¹⁹ Based on the work of the Earth Science Information Partners (ESIP). For more guidance on how to differentiate between major and minor versions, see:

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
16	Rights	0-n	Any rights information for this resource.	Free text. *** Provide a rights management statement for the resource or reference a service providing such information. Include embargo information if applicable. Use the complete title of a license and include version information if applicable. Example: Creative Commons Attribution 3.0 Germany License
16.1	rightsURI	0-1	The URI of the license.	Example: http://creativecommons.org/licenses/by/3.0/de/deed.en
17	Description	0-n	All additional information that does not fit in any of the other categories. May be used for technical information.	The format is open *** It is a best practice to supply a description.
17.1	descriptionType	1	The type of the Description.	If Description is used, descriptionType is mandatory. <i>Controlled List Values:</i> Abstract Methods SeriesInformation TableOfContents Other See appendix for definitions.
18	GeoLocation	0-n	Spatial region or named place where the data was gathered or about which the data is focused.	Repeat this property to indicate several different locations.

<i>ID</i>	<i>DataCite-Property</i>	<i>Occ</i>	<i>Definition</i>	<i>Allowed values, examples, other constraints</i>
18.1	geoLocationPoint	0-1	A point location in space	A point contains a single latitude-longitude pair, separated by whitespace. Detailed usage instructions ²⁰ Example: <geoLocationPoint>31.233 - 67.302 </geoLocationPoint>
18.2	geoLocationBox	0-1	The spatial limits of a place.	A box contains two white space separated latitude-longitude pairs, with each pair separated by whitespace. The first pair is the lower corner, the second is the upper corner. Detailed usage instructions ²⁰ Example: <geoLocationBox>42.893-71.032 41.090-68.211</geoLocationBox>
18.3	geoLocationPlace	0-1	Description of a geographic location	Free text. Use to describe a geographic location.

3 XML Example

Examples for various resource types can be found at <http://schema.datacite.org/meta/kernel-3.0/index.html>¹⁹.

4 XML Schema

The XML Schema is available here:

<http://schema.datacite.org/meta/kernel-3.0/metadata.xsd>

doi:10.5438/0009

Note that the schema and this documentation will always have the same version number.

Each subsequent version of the schema will be at this same location using an address composed in the same manner, that is: <http://schema.datacite.org/meta/kernel-versionnumber/metadata.xsd>.

Earlier versions will continue to be available at their previous locations for backward compatibility.

²⁰ Use WGS 84 (World Geodetic System) coordinates. Use only decimal numbers for coordinates. Longitudes are -180 to 180 (0 is Greenwich, negative numbers are west, positive numbers are east), Latitudes are -90 to 90 (0 is the equator; negative numbers are south, positive numbers north).

5 Other DataCite Services

For information about other DataCite services that pertain to DataCite metadata records, including the Metadata Store, Metadata Search and Content Negotiation, please see DataCite's Service page at <http://www.datacite.org/services>.

Appendices

Appendix 1: Controlled List Definitions

In Appendix 1, as in Sections 2.1 and 2.3 above, controlled list values that enhance the prospect that the resource's metadata will be found, cited and linked are indicated by shading.

contributorType

Table 5: Description of contributorType

<i>Option</i>	<i>Description</i>	<i>Usage Notes</i>
ContactPerson	Person with knowledge of how to access, troubleshoot, or otherwise field issues related to the resource	May also be "Point of Contact" in organization that controls access to the resource, if that organization is different from Publisher, Distributor, Data Manager
DataCollector	Person/institution responsible for finding, gathering/collecting data under the guidelines of the author(s) or Principal Investigator (PI)	May also use when crediting survey conductors, interviewers, event or condition observers, person responsible for monitoring key instrument data.
DataManager	Person (or organization with a staff of data managers, such as a data center) responsible for maintaining the finished resource.	The work done by this person or organization ensures that the resource is periodically "refreshed" in terms of software/hardware support, is kept available or is protected from unauthorized access, is stored in accordance with industry standards, and is handled in accordance with the records management requirements applicable to it.
Distributor	Institution tasked with responsibility to generate/disseminate copies of the resource in either electronic or print form.	Works stored in more than one archive/repository may credit each as a distributor.
Editor	A person who oversees the details related to the publication format of the resource.	
Funder	Institution that provided financial support for the development of the resource.	Recommended for discovery. Includes organizations that provide funding via regular budget allocations, through grants or awards

<i>Option</i>	<i>Description</i>	<i>Usage Notes</i>
HostingInstitution	Typically, the organization allowing the resource to be available on the Internet through the provision of its hardware/software/operating support.	May also be used for an organization that stores the data offline. Often a data center (if that data center is not the “publisher” of the resource).
Producer	Typically a person or organization responsible for the artistry and form of a media product.	In the data industry, this may be a company “producing” DVDs that package data for future dissemination by a distributor.
ProjectLeader	Person officially designated as head of project team or sub-project team instrumental in the work necessary to development of the resource.	The Project Leader is not “removed” from the work that resulted in the resource; he or she remains intimately involved throughout the life of the particular project team.
ProjectManager	Person officially designated as manager of a project. Project may consist of one or many project teams and sub-teams.	The manager of a Project normally has more administrative responsibility than actual work involvement.
ProjectMember	Person on the membership list of a designated project/project team	This vocabulary may or may not indicate the quality, quantity, or substance of the person’s involvement.
RegistrationAgency	Institution/organization officially appointed by a Registration Authority to handle specific tasks within a defined area of responsibility	DataCite is a Registration Agency for the International DOI Foundation (IDF). One of DataCite’s tasks is to assign DOI prefixes to the allocating agents who then assign the full, specific character string to data clients, provide metadata back to the DataCite registry, etc.
RegistrationAuthority	A standards-setting body from which Registration Agencies obtain official recognition and guidance.	The IDF serves as the Registration Authority for the International Standards Organization (ISO) in the area/domain of Digital Object Identifiers.
RelatedPerson	A person without a specifically defined role in the development of the resource, but who is someone the author wishes to recognize.	This person could be an author’s intellectual mentor, a person providing intellectual leadership in the discipline or subject domain, etc.

<i>Option</i>	<i>Description</i>	<i>Usage Notes</i>
Researcher	A person involved in analyzing data or the results of an experiment or formal study. May indicate an intern or assistant to one of the authors who helped with research but who was not so “key” as to be listed as an author.	Should be a person, not an institution. Note that a person involved in the gathering of data would fall under the contributorType “DataCollector.” The researcher may find additional data online and correlate it to the data collected for the experiment or study, for example.
ResearchGroup	Typically refers to a group of individuals with a lab, department, or division; the group has a particular, defined focus of activity.	May operate at a narrower level of scope; may or may not hold less administrative responsibility than a project team.
RightsHolder	Person or institution owning or managing property rights, including intellectual property rights over the resource.	
Sponsor	Person or organization that issued a contract or under the auspices of which a work has been written, printed, published, developed, etc.	Includes organizations that provide in-kind support, through donation, provision of people or a facility or instrumentation necessary for the development of the resource, etc.
Supervisor	Designated administrator over one or more groups/teams working to produce a resource or over one or more steps of a development process.	
WorkPackageLeader	A Work Package is a recognized data product, not all of which is included in publication. The package, instead, may include notes, discarded documents, etc. The Work Package Leader is responsible for ensuring the comprehensive contents, versioning, and availability of the Work Package during the development of the resource.	
Other	Any person or institution making a significant contribution to the development and/or maintenance of the resource, but whose contribution does not “fit” other controlled vocabulary for contributorType.	Could be a photographer, artist, or writer whose contribution helped to publicize the resource (as opposed to creating it), a reviewer of the resource, someone providing administrative services to the author (such as depositing updates into an online repository, analysing usage, etc.), or one of many other roles.

dateType

NOTE: To indicate a date range, follow the RKMS-ISO8601 standard for depicting date ranges.

For example:

```
<date dateType="created">2012-03-01/2012-03-05</date>
```

Table 6: Description of dateType

<i>Option</i>	<i>Description</i>	<i>Usage Notes</i>
Accepted	The date that the publisher accepted the resource into their system.	To indicate the start of an embargo period, use Submitted or Accepted, as appropriate.
Available	The date the resource is made publicly available. May be a range.	To indicate the end of an embargo period, use Available.
Copyrighted	The specific, documented date at which the resource receives a copyrighted status, if applicable.	
Collected	The date or date range in which the resource content was collected.	To indicate precise or particular timeframes in which research was conducted.
Created	The date the resource itself was put together; this could be a date range or a single date for a final component, e.g., the finalised file with all of the data.	Recommended for discovery.
Issued	The date that the resource is published or distributed e.g. to a data center	
Submitted	The date the creator submits the resource to the publisher. This could be different from Accepted if the publisher then applies a selection process.	Recommended for discovery. To indicate the start of an embargo period, use Submitted or Accepted, as appropriate.
Updated	The date of the last update to the resource, when the resource is being added to. May be a range.	
Valid	The date or date range during which the dataset or resource is accurate. May be a range.	

resourceTypeGeneral

Table 7: Description of resourceTypeGeneral

<i>Option</i>	<i>Description²¹</i>	<i>Examples and Usage Notes</i>	<i>Suggested Dublin Core Mapping</i>
Audiovisual	A series of visual representations imparting an impression of motion when shown in succession. May or may not include sound.	May be used for films, video, etc, Ex: http://data.datacite.org/10.7794/USR3225/ARC.F/0000001	MovingImage
Collection	An aggregation of resources of various types. If a collection exists of a single type, use the single type to describe it.	A collection of samples, or various files making up a report. Ex: http://data.datacite.org/10.5284/1001038	Collection
Dataset	Data encoded in a defined structure.	Data file or files, Ex: http://data.datacite.org/10.4231/D39Z90B9T	Dataset
Event	A non-persistent, time-based occurrence.	Descriptive information and/or content that is the basis for discovery of the purpose, location, duration, and responsible agents associated with an event such as a webcast or convention.. Ex: http://data.datacite.org/10.7269/P3RN35SZ	Event
Image	A visual representation other than text.	Digitized or born digital images, drawings or photographs. Ex: http://data.datacite.org/10.5060/D4RN35SD	Image, StillImage
InteractiveResource	A resource requiring interaction from the user to be understood, executed, or experienced	Training modules, files that require use of a viewer (e.g., Flash), or query/response portals. Ex: http://data.datacite.org/10.7269/P3TB14TR	InteractiveResource

²¹Where there is direct correspondence with the Dublin Core Metadata, DataCite definitions have borrowed liberally from the DCMI definitions. See: <http://dublincore.org/documents/dcmi-terms/#terms-DCMIType>

Option	Description²¹	Examples and Usage Notes	Suggested Dublin Core Mapping
Model	An abstract, conceptual, graphical, mathematical or visualization model that represents empirical objects, phenomena, or physical processes.	Modelled descriptions of, for example, different aspects of languages or a molecular biology reaction chain.	N/A
PhysicalObject	An inanimate, three-dimensional object or substance.	Artifacts, specimens. Ex: http://data.datacite.org/10.7299/X78052RB	PhysicalObject
Service	A system that provides one or more functions of value to the end-user.	Data management service, authentication service, or photocopying service.	Service
Software	A computer program in source code (text) or compiled form.	Software supporting research. Ex: http://data.datacite.org/10.5524/100046	Software
Sound	A resource primarily intended to be heard.	Audio recording. Ex: http://data.datacite.org/10.4231/3DZVR.1288284313124	Sound
Text	A resource consisting primarily of words for reading.	Grey literature, lab notes, accompanying materials. Ex: http://data.datacite.org/10.5682/9786065914018	Text
Workflow	A structured series of steps which can be executed to produce a final outcome, allowing users a means to specify and enact their work in a more reproducible manner.	Computational workflows involving sequential operations made on data by wrapped software and may be specified in a format belonging to a workflow management system, such as Taverna (http://www.taverna.org.uk/). More. ²²	N/A
Other	If selected, supply a value for ResourceType.		

²² An education module on workflows prepared by DataONE is available at http://www.dataone.org/sites/all/documents/L10_AnalysisWorkflows.pptx

relatedIdentifierType

Table 8: Description of relatedIdentifierType

Option	Full Name	Example
ARK	Archival Resource Key; URL designed to support long-term access to information objects. In general, ARK syntax is of the form (brackets indicate [optional] elements: [http://NMA/]ark:/NAAN/Name [Qualifier])	<code><relatedIdentifier relatedIdentifierType="ARK" relationType="IsCitedBy">ark:/13030/tqb3kh97gh8w </relatedIdentifier></code>
DOI	Digital Object Identifier; a character string used to uniquely identify an object. A DOI name is divided into two parts, a prefix and a suffix, separated by a slash.	<code><relatedIdentifier relatedIdentifierType="DOI" relationType="IsSupplementTo"> 10.1016/j.epsl.2011.11.037 </relatedIdentifier></code>
EAN13	European Article Number, now renamed International Article Number, but retaining the original acronym, is a 13-digit barcoding standard which is a superset of the original 12-digit Universal Product Code (UPC) system.	<code><relatedIdentifier relatedIdentifierType="EAN13" relationType="Cites">9783468111242 </relatedIdentifier></code>
EISSN	Electronic International Standard Serial Number; ISSN used to identify periodicals in electronic form (eISSN or e-ISSN)	<code><relatedIdentifier relatedIdentifierType="eISSN" relationType="Cites">1562-6865 </relatedIdentifier></code>
Handle	A handle is an abstract reference to a resource.	<code><relatedIdentifier relatedIdentifierType="Handle" relationType="References">10013/epic.10033 </relatedIdentifier></code>
ISBN	International Standard Book Number; a unique numeric book identifier. There are 2 formats: a 10-digit ISBN format and a 13-digit ISBN.	<code><relatedIdentifier><relatedIdentifier relatedIdentifierType="ISBN" relationType="IsPartOf">978-3-905673-82-1 </relatedIdentifier></code>
ISSN	International Standard Serial Number; a unique 8-digit number used to identify a print or electronic periodical publication.	<code><relatedIdentifier relatedIdentifierType="ISSN" relationType="IsPartOf">0077-5606 </relatedIdentifier></code>

<i>Option</i>	<i>Full Name</i>	<i>Example</i>
ISTC	International Standard Text Code; a unique “number” assigned to a textual work. An ISTC consists of 16 numbers and/or letters.	<code><relatedIdentifier relatedIdentifierType="ISTC" relationType="Cites">0A9 2002 12B4A105 7 </relatedIdentifier></code>
LISSN	The linking ISSN or ISSN-L enables collocation or linking among different media versions of a continuing resource.	<code><relatedIdentifier relatedIdentifierType="LISSN" relationType="Cites">1188-1534</relatedIdentifier></code>
LSID	Life Science Identifiers; a unique identifier for data in the Life Science domain. Format: urn:lsid:authority:namespace:identifier:revision	<code><relatedIdentifier relatedIdentifierType="LSID" relationType="Cites"> urn:lsid:ubio.org:namebank:11815</relatedIdentifier></code>
PMID	PubMed identifier; a unique number assigned to each PubMed record.	<code><relatedIdentifier relatedIdentifierType="PMID" relationType="IsReferencedBy">12082125</relatedIdentifier></code>
PURL	Persistent Uniform Resource Locator. A PURL has three parts: (1) a <i>protocol</i> , (2) a <i>resolver address</i> , and (3) a <i>name</i> .	<code><relatedIdentifier relatedIdentifierType="PURL" relationType="Cites"> http://purl.oclc.org/foo/bar</relatedIdentifier></code>
UPC	Universal Product Code is a barcode symbology used for tracking trade items in stores. Its most common form, the UPC-A, consists of 12 numerical digits.	<code><relatedIdentifier relatedIdentifierType="UPC" relationType="Cites"> 123456789999</relatedIdentifier></code>
URL	Uniform Resource Locator, also known as web address, is a specific character string that constitutes a reference to a resource. The syntax is: scheme://domain:port/path?query_string#fragment_id	<code><relatedIdentifier relatedIdentifierType="URL" relationType="IsCitedBy">http://www.heatflow.und.edu/ index2.html</relatedIdentifier></code>
URN	Uniform Resource Name; is a unique and persistent identifier of an electronic document. The syntax is: urn:<NID>:<NSS> The leading urn: sequence is case-insensitive, <NID> is the namespace identifier, <NSS> is the namespace-specific string.	<code><relatedIdentifier relatedIdentifierType="URN" relationType="IsSupplementTo">urn:nbn:de:101:1- 201102033592</relatedIdentifier></code>

relationType

Description of the relationship of the resource being registered (A) and the related resource (B).

Table 9: Description of relationType

Option	Definition	Example and Usage Notes
IsCitedBy	indicates that B includes A in a citation	Recommended for discovery. <code><relatedIdentifier relatedIdentifierType="DOI" relationType="IsCitedBy">10.4232/10.ASEAS-5.2-1</relatedIdentifier></code>
Cites	indicates that A includes B in a citation	Recommended for discovery. <code><relatedIdentifier relatedIdentifierType="ISBN" relationType="Cites">0761964312</relatedIdentifier></code>
IsSupplementTo	indicates that A is a supplement to B	Recommended for discovery. <code><relatedIdentifier relatedIdentifierType="URN" relationType="IsSupplementTo">http://nbn-resolving.de/urn:nbn:de:0168-ssoar-13172</relatedIdentifier></code>
IsSupplementedBy	indicates that B is a supplement to A	Recommended for discovery. <code><relatedIdentifier relatedIdentifierType="PMID" relationType="IsSupplementedBy">16911322/</relatedIdentifier></code>
IsContinuedBy	indicates A is continued by the work B	<code><relatedIdentifier relatedIdentifierType="URN" relationType="IsContinuedBy">http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-4967</relatedIdentifier></code>
Continues	indicates A is a continuation of the work B	<code><relatedIdentifier relatedIdentifierType="URN" relationType="Continues">http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-4966</relatedIdentifier></code>
HasMetadata	indicates resource A has additional metadata B	<code><relatedIdentifier relatedIdentifierType="DOI" relationType="HasMetadata">10.1234/567890 relatedMetadataScheme="DDI-L" schemeURI="http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/XMLSchema/instance.xsd"</relatedIdentifier></code>
IsMetadataFor	indicates additional metadata A for a resource B	<code><relatedIdentifier relatedIdentifierType="DOI" relationType="IsMetadataFor">10.1234/567891 relatedMetadataScheme="DDI-L" schemeURI="http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/XMLSchema/instance.xsd"</relatedIdentifier></code>

Option	Definition	Example and Usage Notes
IsNewVersionOf	indicates A is a new edition of B, where the new edition has been modified or updated	<code><relatedIdentifier relatedIdentifierType="DOI" relationType="IsNewVersionOf">10.5438/0005</relatedIdentifier></code>
IsPreviousVersionOf	indicates A is a previous edition of B	<code><relatedIdentifier relatedIdentifierType="DOI" relationType="IsPreviousVersionOf">10.5438/0007</relatedIdentifier></code>
IsPartOf	indicates A is a portion of B; may be used for elements of a series	Recommended for discovery. <code><relatedIdentifier relatedIdentifierType="ISBN" relationType="IsPartOf">0-486-27557-4</relatedIdentifier></code>
HasPart	indicates A includes the part B	Recommended for discovery. <code><relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">10.1234/7894</relatedIdentifier></code>
IsReferencedBy	indicates A is used as a source of information by B	<code><relatedIdentifier relatedIdentifierType="URL" relationType="IsReferencedBy">http://www.testpubl.de</relatedIdentifier></code>
References	indicates B is used as a source of information for A	<code><relatedIdentifier relatedIdentifierType="URN" relationType="References">http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-963</relatedIdentifier></code>
IsDocumentedBy	indicates B is documentation about/explaining A)	<code><relatedIdentifier relatedIdentifierType="URL" relationType="IsDocumentedBy">http://tobias-lib.uni-tuebingen.de/volltexte/2000/96/</relatedIdentifier></code>
Documents	indicates A is documentation about/explaining B	<code><relatedIdentifier relatedIdentifierType="DOI" relationType="Documents">10.1234/7836</relatedIdentifier></code>
isCompiledBy	indicates B is used to compile or create A	<code><relatedIdentifier relatedIdentifierType="URL" relationType="isCompiledBy">http://d-nb.info/gnd/4513749-3</relatedIdentifier></code>
Compiles	indicates B is the result of a compile or creation event using A	<code><relatedIdentifier relatedIdentifierType="URN" relationType="Compiles">http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-963</relatedIdentifier></code>

Option	Definition	Example and Usage Notes
IsVariantFormOf	indicates A is a variant or different form of B, e.g. calculated or calibrated form or different packaging	<pre><relatedIdentifier relatedIdentifierType="DOI" relationType="IsVariantFormOf">10.1234/8675 </relatedIdentifier></pre>
IsOriginalFormOf	indicates A is the original form of B	<pre><relatedIdentifier relatedIdentifierType="DOI" relationType="IsOriginalFormOf">10.1234/9035 </relatedIdentifier></pre>
IsIdenticalTo	indicates that A is identical to B, for use when there is a need to register two separate instances of the same resource	<pre><relatedIdentifier relatedIdentifierType="URL" relationType="IsIdenticalTo">http://oac.cdlib.org/findaid/ark:/13030/c8r78fzq </relatedIdentifier></pre> <p>IsIdenticalTo should be used for a resource that is the same as the registered resource but is saved on another location, maybe another institution.</p>

descriptionType

Table 10: Description of contributorType

Option	Definition	Usage Notes
Abstract	A brief description of the resource and the context in which the resource was created.	Recommended for discovery. Use " " to indicate a line break for improved rendering of multiple paragraphs, but otherwise no html markup. Example: http://data.datacite.org/10.1594/PANGAEA.771774
Methods	The methodology employed for the study or research.	Recommended for discovery. For example, see section "Sampling, Processing and Quality Control Methods" in the following dataset record: http://kn.b.ecoinformatics.org/knb/metacat?action=read&qformat=kn&sessionid=0&docid=knb-lter-gce.275.16 .
SeriesInformation	Information about a repeating series, such as volume, issue, number.	For use with grey literature. If providing an ISSN, use property 12 (RelatedIdentifier), relatedIdentifierType=ISSN. For dataset series, use property 12 (RelatedIdentifier) and describe the relationships with isPartOf or HasPart. Example: http://data.datacite.org/10.4229/23RDEUPVSEC2008-5CO.8.3
TableOfContents	A listing of the Table of Contents.	Use " " to indicate a line break for improved rendering of multiple paragraphs, but otherwise no html markup. Example: http://data.datacite.org/10.5678/LCRS/FOR816.CIT.1031
Other	Other description information that does not fit into an existing category.	Use for any other description type.

Appendix 2: Earlier Version Update Notes

Version 2.1 Update

Version 2.1 of the DataCite Metadata Schema introduced several changes, as noted below:

- Addition of a namespace (<http://datacite.org/schema/namespace>) to the schema in order to support OAI PMH compatibility
- Enforcement of content for mandatory properties
- New type for the `Date` property to conform with the specification that it handles both YYYY and YYYY-MM-DD values

Version 2.1 of the DataCite Metadata Schema documentation includes these changes:

- Addition of a column to the Mandatory and Optional Properties tables providing an indicator of whether the property being described is an attribute or a child of the corresponding property that has preceded it
- Revision of the allowed values description for the attribute 12.2 `relationType`. These have been reviewed and rewritten for increased clarity. In several cases, corrections to the definitions occurred.

Version 2.2 Update

Version 2.2 of the DataCite Metadata Schema introduced several changes, as noted below:

- Addition of “URL” to list of allowed values for `relatedIdentifierType`
- Addition of the following values to list of allowed values for `contributorType`: Producer, Distributor, RelatedPerson, Supervisor, Sponsor, Funder, RightsHolder
- Addition of “SeriesInformation” to list of allowed values for `descriptionType`
- Addition of “Model” to list of allowed values for `resourceTypeGeneral`

Version 2.2 of the DataCite Metadata Schema documentation included these changes:

- Provision of more examples of xml for different types of resources
- Explanation of the `PublicationYear` property in consideration of the requirements of citation. A change to the definition of the `Publisher` property, which now reads, “The name of the entity that holds, archives, publishes, prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role.”