Metadata Schema for the Persistent Identification of Instruments

RDA Recommendation

DOI: 10.15497/RDA00070

Authors: Rolf Krahl, Louise Darroch, Robert Huber, Anusuriya Devaraju, Jens Klump, Ted Habermann, Markus Stocker, RDA PIDINST WG members

Published: 30 March 2022

Version: 1.0

Abstract: Instruments play an essential role in creating research data but they are often only identified in scientific literature using free text. Through the use of globally unique, persistent identifiers (PIDs) such as the Digital Object Identifier (DOI), it is now common practice to establish traceable links between scientific literature and the data that generated it. Such cross-linking has received considerable attention in the community in recent years and has been generalized to other entities, including people, organizations, funders, etc. Given the importance of instruments and associated metadata to the assessment of data quality and data reuse, globally unique, persistent and resolvable identification of instruments is also crucial. The Research Data Alliance Persistent Identification of Instruments Working Group (PIDINST WG) developed and demonstrated a solution for publishing persistent identifiers for instruments at established PID providers. This solution comprises a metadata schema for the persistent identification of instruments (PIDINST Schema).

Impact: This PIDINST Schema is the recommended schema for the persistent identification of instruments by institutions or bodies using or managing instruments. The Schema is PID provider independent and has been tested with ePIC and DataCite by numerous institutions including Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, British Oceanographic Data Centre, EUDAT.

Language: English

RDA webpage: https://www.rd-alliance.org/group/persistent-identification-instruments-wg/outcomes/metadata-schema-persistent-identification

Citation and Download: Krahl, R., Darroch, L., Huber, R., Devaraju, A., Klump, J., Habermann, T., Stocker, M., & The Research Data Alliance Persistent Identification of Instruments Working Group members. (2021). Metadata Schema for the Persistent Identification of Instruments. *Research Data Alliance*. https://doi.org/10.15497/RDA00070

Metadata Schema for the Persistent Identification of Instruments

RDA Persistent Identification of Instruments WG

28 March 2022

Version	1.0
Date	28 March 2022

Contributors

- *Rolf Krahl* (rolf.krahl@helmholtz-berlin.de, https://orcid.org/0000-0002-1266-3819), Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Str. 15, 12489 Berlin, Germany
- *Louise Darroch* (louise.darroch@bodc.ac.uk, https://orcid.org/0000-0003-4163-9575), British Oceanographic Data Centre, National Oceanography Centre, Liverpool, L3 5DA, United Kingdom
- *Robert Huber* (rhuber@uni-bremen.de, https://orcid.org/0000-0003-3000-0020), MARUM - Center for Marine Environmental Sciences, University of Bremen, Leobener Str. 8, 28359 Bremen, Germany
- *Anusuriya Devaraju* (a.devaraju@uq.edu.au, https://orcid.org/0000-0003-0870-3192), Terrestrial Ecosystem Research Network (TERN), The University of Queensland, Long Pocket Precinct, Level 5 Foxtail Building #1019, 80 Meiers Road, Indooroopilly QLD 4068, Australia
- *Jens Klump* (jens.klump@csiro.au, https://orcid.org/0000-0001-5911-6022), CSIRO Mineral Resources, 26 Dick Perry Avenue, Kensington WA 6151, Australia
- *Ted Habermann* (ted@metadatagamechangers.com, http://orcid.org/0000-0003-3585-6733), Metadata Game Changers, Boulder, Colorado, USA 80304
- *Markus Stocker* (markus.stocker@tib.eu, https://orcid.org/0000-0001-5492-3212), TIB – Leibniz Information Centre for Science and Technology, Welfengarten 1 B, 30167 Hannover, Germany and Leibniz University Hannover, Welfengarten 1, 30167 Hannover, Germany
- *The Research Data Alliance Persistent Identification of Instruments Working Group members* (https://www.rd-alliance.org/node/57186/members)

Contents

1. Introduction

Instruments, such as sensors used in environmental science, DNA sequencers used in life sciences or laboratory engines used for medical domains, are widespread in most fields of applied sciences. The ability to link a physical instrument to the data that it generated and contextual metadata such as when, where and how it was operated, is critical for the accurate interpretation of that data.

The Persistent Identification of Instruments Working Group (PIDINST) is a working group of the Research Data Alliance (RDA). It aims to establish a cross-discipline, practical solution for the globally unique and persistent identification of instruments operated in the sciences that can act as an anchor for associated metadata. The group considers instrument instances, e.g. the individual physical objects, as opposed to instrument types or models. See [1] for an overview of the work developed by the RDA PIDINST WG.

The solution uses a resolvable persistent identifier (PID) published with descriptive metadata that provides enough information to identify the instrument. Instrument PIDs also provide more context to the instrument itself by linking other instruments and platforms, descriptive information and other external metadata about the instrument, and manufacturer and owner of the instrument. This information can then be automatically aggregated following the so-called PID Graph [2].

This document defines the metadata schema that PIDINST developed for the description of an instrument instance. Institutions or organizations operating scientific instruments may use the schema when registering the metadata describing their instrument along with the persistent identifier through a relevant PID infrastructure. The schema is publicly available for reuse and maintained in the code repository of the working group [3].

2. Metadata Schema

The PIDINST metadata schema provides core properties to represent instrument descriptions. It contains essential information to identify the individual instrument instance and to link the instance with other sources of information and related entities.

The Working Group strives for a metadata schema that is applicable to various kinds of instruments being used in all scientific domains. For this reason, we kept the schema generic. Technical details such as specifications, data sheets, or detailed technique descriptions were not included as these may only apply to a subclass of instruments relevant to a particular scientific subdomain. The core metadata of an instrument may be extended with more detailed, community-specific descriptions by linking the PIDINST record with supplementary metadata using the properties RelatedIdentifier with relationType IsDescribedBy or HasMetadata provided in the schema.

2.1. Overview

Table 1 provides an overview of the properties in the PIDINST schema. The last column indicates whether the property is:

- Mandatory (M): the property must be provided for the metadata record to be valid,
- Recommended (R): the property is optional, but it is recommended to be added if applicable, or
- Optional (O): the property may be added to provide a richer description.

In the present version of the schema, all properties are either mandatory or recommended. It is envisioned that additional optional properties (in addition to mandatory and recommended properties) will be added in the future.

ID	*Property*	*Obligation*
1	Identifier	M
2	SchemaVersion	M
3	LandingPage	M
4	Name	M
5	Owner	M
6	Manufacturer	M
7	Model	R
8	Description	R
9	InstrumentType	R
10	MeasuredVariable	R
11	Date	R
12	RelatedIdentifier	R
13	AlternateIdentifier	R

Table 1: Overview of PIDINST properties

If a value for any of the mandatory properties is not available, one of the standard values for unknown information may be used instead, see Appendix A.

2.2. Definition of the PIDINST properties

Table 2 provides a detailed description for the properties in the PIDINST metadata schema. Simple properties, such as Name and Description, take just one value. Complex properties are composed of multiple sub-properties. For instance, Owner is composed of ownerName, ownerContact, and ownerIdentifier, the latter in turn having another sub-property ownerIdentifierType. The structure of properties and sub-properties is indicated in the numbering in the ID column, e.g. the property 5.3 (ownerIdentifier) has a sub-property 5.3.1 (ownerIdentifierType).

The number of occurrences for any property in one PIDINST metadata record is indicated in the column labeled "Occ". The meaning is:

- 1: a mandatory property that must appear once,
- 0–1: an optional property that may appear at most once,
- 1–n: a mandatory property with multiple values, e.g. the property may appear one or more times in a single record,
- 0–n: a multivalued optional property that may appear zero or more times in a record.

In the case of a sub-property, the occurrence is to be understood within each of the parent entries. For example, a PIDINST record may have zero or more RelatedIdentifier (occ. 0–n).

Each of them must have one relatedIdentifierType (occ. 1) and one relationType (occ. 1) and may optionally have one relatedIdentifierName (occ. 0–1). Another example: a PIDINST record must have one or more Manufacturer (occ. 1–n). Each of them must have a manufacturerName (occ. 1) and may have a manufacturerIdentifier (occ. 0–1). If a manufacturerIdentifier is included for a Manufacturer, it must also have a manufacturerIdentifierType (occ. 1).

ID	*Property*	*Occ*	*Definition*	*Allowed values, constraints, remarks*
1	Identifier	1	Unique string that identifies the instrument instance	
1.1	identifierType	1	Type of the identifier	The type of the identifier depends on the provider being used to register it. In the case of ePIC, the value would be "Handle".
2	SchemaVersion	1	Version number of the PIDINST schema used in this record	Fixed value: "1.0"
3	LandingPage	1	A landing page that the identifier resolves to	URL
4	Name	1	Name by which the instrument instance is known	Free text
5	Owner	1–n	Institution(s) responsible for the management of the instrument. This may include the legal owner, the operator, or an institute providing access to the instrument.	
5.1	ownerName	1	Full name of the owner	Free text
5.2	ownerContact	0–1	Contact address of the owner	Electronic mail address
5.3	ownerIdentifier	0–1	Identifier used to identify the owner	Free text, should be a globally unique identifier
5.3.1	ownerIdentifierType	1	Type of the identifier	Free text, see note below

ID	*Property*	*Occ*	*Definition*	*Allowed values, constraints, remarks*
6	Manufacturer	1–n	The instrument's manufacturer(s) or developer. This may also be the owner for custom-build instruments.	
6.1	manufacturerName	1	Full name of the manufacturer	Free text
6.2	manufacturerIdentifier	0–1	Identifier used to identify the manufacturer	Free text, should be a globally unique identifier
6.2.1	manufacturerIdentifierType	1	Type of the identifier	Free text, see note below
7	Model	0–1	Name of the model or type of device as attributed by the manufacturer	
7.1	modelName	1	Full name of the model	Free text
7.2	modelIdentifier	0–1	Identifier used to identify the model	Free text, should be a globally unique identifier
7.2.1	modelIdentifierType	1	Type of the identifier	Free text, see note below
8	Description	0–1	Technical description of the device and its capabilities	Free text
9	InstrumentType	0–n	Classification of the type of the instrument	
9.1	instrumentTypeName	1	Full name of the instrument type	Free text, see note below
9.2	instrumentTypeIdentifier	0–1	Identifier used to identify the type of the instrument	Free text, should be a globally unique identifier
9.2.1	instrumentTypeIdentifierType	1	Type of the identifier	Free text, see note below
10	MeasuredVariable	0–n	The variable(s) that this instrument measures or observes	Free text
11	Date	0–n	Dates relevant to the instrument	ISO 8601

ID	*Property*	*Occ*	*Definition*	*Allowed values, constraints, remarks*
11.1	dateType	1	The type of the date	Controlled list of values: • Commissioned • DeCommissioned
12	RelatedIdentifier	0–n	Identifiers of related resources	Free text, must be globally unique identifiers
12.1	relatedIdentifierType	1	Type of the identifier	Controlled list of values: • ARK • arXiv • bibcode • DOI • EAN13 • EISSN • Handle • IGSN • ISBN • ISSN • ISTC • LISSN • PMID • PURL • RAiD • RRID • UPC • URL • URN • w3id
12.2	relationType	1	Description of the relationship	Controlled list of values: • IsDescribedBy • IsNewVersionOf • IsPreviousVersionOf • HasComponent • IsComponentOf • References • HasMetadata • WasUsedIn • IsIdenticalTo • IsAttachedTo
12.3	relatedIdentifierName	0–1	A name for the related resource, may be used to give a hint on the content of that resource	Free text

ID	*Property*	*Occ*	*Definition*	*Allowed values, constraints, remarks*
13	AlternateIdentifier	0–n	Identifiers other than the PIDINST pertaining to the same instrument instance. This should be used if the instrument has a serial number. Other possible uses include an owner's inventory number or an entry in some instrument data base.	Free text, should be unique identifiers
13.1	alternateIdentifierType	1	Type of the identifier	Controlled list of values: • SerialNumber • InventoryNumber • Other
13.2	alternateIdentifierName	0–1	A supplementary name for the identifier type. This is mostly useful if alternateIdentifierType is Other.	Free text

Table 2: Expanded PIDINST properties

There is not one single comprehensive classification for instrument types that would be applicable to all kinds of instruments being used in all scientific domains. That is why we can't prescribe a controlled list of values for instrumentTypeName, but allow free text. There are however controlled vocabularies or ontologies to classify instrument types for some scientific domains. If the instrument is used in a domain that has a well established instrument type classification, it is recommended to use these terms in instrumentTypeName. If furthermore the classification defines identifiers for the terms, it is recommended to link them using instrumentTypeIdentifier.

It is important that links to related resources can be resolved in a reliable way. Therefore, well established persistent identifier schemes should preferably be used to make these links and that is why we prescribe a controlled list of values for relatedIdentifierType. For ownerIdentifierType, manufacturerIdentifierType, modelIdentifierType, and instrumentTypeIdentifierType we anticipate that there might be other, less well established identifier types that may need to be used in some cases. Hence we allow free text. Nevertheless, the values defined for relatedIdentifierType should also be used here, whenever it is possible.

As formalized controlled vocabularies evolve and become common standard across domains, future versions of this schema may adopt them and change properties that are free text in the current version to use a controlled list of values instead.

2.3. Definition of terms in controlled lists of values

For some properties, the allowed values are constrained to a controlled list. In the following, we define the semantics of the terms and provide guidance on when to use them.

2.3.1. dateType

The Date property allows references to relevant events in the lifetime of the instrument. The dateType sub-property indicates the nature of that event. Table 3 lists the allowed values for dateType.

Value	*Definition*
Commissioned	The date when the instrument started to be in operation
DeCommissioned	The date when the instrument ceased to be in operation

Table 3: Values for dateType

2.3.2. relatedIdentifierType

The RelatedIdentifier property should be used to link other resources of any kind related to the instrument, referring to a persistent identifier of this resource. The relatedIdentifierType sub-property indicates the type of persistent identifier being used in the link. The allowed values for relatedIdentifierType are listed in Table 4. Most of these values have been adopted from the DataCite Metadata Schema [4].

Value	*Definition*
ARK	Archival Resource Key: a URI designed to support long-term access to information objects.
arXiv	arXiv identifier: arXiv.org is a repository of preprints of scientific papers in the fields of mathematics, physics, astronomy, computer science, quantitative biology, statistics, and quantitative finance.
bibcode	Astrophysics Data System bibliographic codes: a standardized 19-character identifier.
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.
EAN13	European Article Number, now renamed International Article Number, but retaining the original acronym, is a 13-digit barcoding standard that is a superset of the original 12-digit Universal Product Code (UPC) system.
EISSN	Electronic International Standard Serial Number: ISSN used to identify periodicals in electronic form (eISSN or e-ISSN).
Handle	This refers specifically to an ID in the Handle system operated by the Corporation for National Research Initiatives (CNRI).

Value	*Definition*
IGSN	A persistent unique identifier for physical samples and specimens. See `https://www.igsn.org/`
ISBN	International Standard Book Number: a unique numeric book identifier.
ISSN	International Standard Serial Number: a unique 8-digit number used to identify a print or electronic periodical publication.
ISTC	International Standard Text Code: a unique "number" assigned to a textual work. An ISTC consists of 16 numbers and/or letters.
LISSN	The linking ISSN or ISSN-L enables collocation or linking among different media versions of a continuing resource.
PMID	PubMed identifier: a unique number assigned to each PubMed record.
PURL	Persistent Uniform Resource Locator.
RAiD	Research Activity Identifier, see `https://www.raid.org.au/`
RRID	Research Resource Identifiers, see `https://www.rrids.org/`
UPC	Universal Product Code is a barcode symbology used for tracking trade items in stores.
URL	Uniform Resource Locator, also known as web address.
URN	Uniform Resource Name: a unique and persistent identifier of an electronic document.
w3id	Permanent identifier for Web applications. Mostly used to publish vocabularies and ontologies.

Table 4: Values for relatedIdentifierType

2.3.3. relationType

The RelatedIdentifier property should be used to link other resources of any kind related to the instrument, referring to a persistent identifier of this resource. The relationType sub-property indicates the nature of that relation. Table 5 lists the allowed values for relationType.

Value	*Definition*
IsDescribedBy	The linked resource is a document describing the instrument.
IsNewVersionOf	If an instrument is substantially modified, a new PID may be attributed to the new version. In that case the old and the new PID should be linked to each other. IsNewVersionOf should be used in the new PID record to link the old instrument before the modification.
IsPreviousVersionOf	If an instrument is substantially modified, a new PID may be attributed to the new version. In that case the old and the new PID should be linked to each other. IsPreviousVersionOf should be used in the old PID record to link the new instrument after the modification.

Value	*Definition*
HasComponent	In the case of a complex instrument, having multiple components that may be considered as instruments in their own right, with their own PIDs, these PIDs should be linked. HasComponent should be used in the PID record of the compound instrument to link the components.
IsComponentOf	In the case of a complex instrument, having multiple components that may be considered as instruments in their own right, with their own PIDs, these PIDs should be linked. IsComponentOf should be used in the PID records of the components to link the compound instrument.
References	This may be used in the generic case, if no other more specific relation type applies.
HasMetadata	If there is additional metadata describing the instrument, possibly using a community specific metadata standard, that metadata record may be linked using HasMetadata.
WasUsedIn	If the instrument has been deployed in some research activity, such as a cruise of a research vessel, WasUsedIn may be used to link that activity.
IsIdenticalTo	If multiple PIDs have been attributed to a single instrument (which should preferably be avoided in the first place), these PID records should be linked to each other using IsIdenticalTo.
IsAttachedTo	If the instrument is permanently attached to another instrument, the PID records for both instruments should link to each other using IsAttachedTo.

Table 5: Values for relationType

2.3.4. alternateIdentifierType

If the instrument is also registered elsewhere using any other kind of identifier, these identifiers may be stored in the AlternateIdentifier property. The alternateIdentifierType sub-property indicates the nature of that identifier. The allowed values are listed in Table 6.

Value	*Definition*
SerialNumber	A serial number attributed by the manufacturer
InventoryNumber	An inventory number used by the owner
Other	Any other kind of identifier

Table 6: Values for alternateIdentifierType

A. Standard values for unknown information

For the case that the value for a mandatory property is unknown or not available, we adopt the standard values for unknown information from the DataCite Metadata Schema [4]. These values are repeated here in Table 7.

Code	*Definition*
:unac	Temporarily inaccessible
:unal	Unallowed, suppressed intentionally
:unap	Not applicable, makes no sense
:unas	Value unassigned (e.g., Untitled)
:unav	Value unavailable, possibly unknown
:unkn	Known to be unknown (e.g., Anonymous, Inconnue)
:none	Never had a value, never will
:null	Explicitly and meaningfully empty
:tba	To be assigned or announced later
:etal	Too numerous to list (et alia)

Table 7: Standard values for unknown information

References

[1] Markus Stocker et al. Persistent Identification of Instruments. *Data Science Journal*, 19(1), 18, 2020. `https://doi.org/10.5334/dsj-2020-018`.

[2] Helena Cousijn et al. Connected Research: The Potential of the PID Graph. *Patterns*, 2(1), 100180, 2021. `https://doi.org/10.1016/j.patter.2020.100180`.

[3] RDA Persistent Identification of Instruments WG. Medadata Schema for the Persistent Identification of Scientific Instruments. `https://github.com/rdawg-pidinst/schema`. Code repository.

[4] DataCite Metadata Working Group. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4, March 2021. `https://doi.org/10.14454/3W3Z-SA82`.