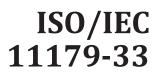# INTERNATIONAL STANDARD

ISO/IEC 11179-33

First edition
2023-01

Information technology — Metadata registries (MDR) —

Part 33:
Metamodel for data set registration

Technologies de l'information — Registres de métadonnées (RM) —

Partie 33: Métamodèle pour l'enregistrement des ensembles de données

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 11179-33:2023
https://standards.iteh.ai/catalog/standards/sist/bb79075f-8b21-4a78-aa84-7d69ff458bf8/iso-iec-
11179-33-2023

COPYRIGHT PROTECTED DOCUMENT

Contents

Page

List of Figures

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 11179-33:2023
https://standards.iteh.ai/catalog/standards/sist/bb79075f-8b21-4a78-aa84-7d69ff458bf8/iso-iec-
11179-33-2023

List of Tables

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 11179-33:2023
https://standards.iteh.ai/catalog/standards/sist/bb79075f-8b21-4a78-aa84-7d69ff458bf8/iso-iec-
11179-33-2023

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see https://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

This first edition of ISO/IEC 11179-33 cancels and replaces ISO/IEC 11179-7:2019, which has been technically revised.

The main changes are as follows.

— The document brings the previously published content into line with the rest of the 4th Edition of ISO/IEC 11179.

— The document also includes enhancements that recognise that a data set can be derived from one or many other data sets.

A list of all parts in the ISO/IEC 11179 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

ISO/IEC 11179-3 specifies the structure of a Metadata registry and provides a metamodel for registry common facilities. This metamodel is intended to be extended by other parts of ISO/IEC 11179 for specific purposes.

This document provides a specification of the extensions to the registry metamodel specified in ISO/IEC 11179-3 to enable the registration of metadata about data sets. These data sets can be part of a government-led open data initiative, or can be data sets that are used within and across organizations for commercial, scientific or academic purposes.

There is a requirement for metadata about these data sets to be readily available to enable the consistent and appropriate use of data and information, and to prevent duplication of work. Having an enhanced Metadata Registry where metadata that describes data sets is registered will facilitate the discovery of appropriate data sets.

The facilities described in this document, together with those described in ISO/IEC 11179-3, ISO/IEC 11179-31 and ISO/IEC 11179-35, provide the ability to record the following information about data sets:

— one or more unique identifiers for the data set;

— the designation or title of the data set;

— a definition or description of the data set that provides sufficient detail to enable a user to quickly understand whether this data set is of interest;

— the date the data set was issued and, if appropriate, the date that subsequent versions of the data set were, or will be, issued;

— the access level and rights associated with the data set;

— the provenance of the data set, i.e. information about the place and time of the origin of the data set, its ownership and the method of the generation of the set;

— a set of keywords or tags that help to explain the data set;

— the language or languages used to describe the data set;

— the temporal and spatial coverages of the data set;

— the accrual periodicity of the data set, i.e., the frequency at which new, revised or updated versions of the data set are made available;

— the details of the distributions of the data set, including the identifier, the title, a description, the media type or file format, the size, the issue date, languages, access level and rights and access and download URLs;

— annotations drawn from a concept system, such as an ontology, to describe the theme or category of the data set or the collection of data sets;

— the details of any contexts, such as a programme, project or business area that use the data set;

— the details of any quality, fitness for role or risk assessments made in respect of the data set;

— any additional descriptions of the data set, including:

— any data elements that are already registered that are included in the data set;

— any information models that describe the structure of the information in the data set;

— any documents which describe aspects of the data set, such as technical information about the data set or developer documentation such as a graphical representation of the data model of the data set;

— the details of any superset/subset hierarchies containing the data set;

— the details of any replacement data set if this data set is superseded;

— the details of any collection of data sets of which this data set is a part, including the identifiers, the designation or title, a definition or description, issue dates, languages, access level, rights, the spatial coverage, the provenance and any quality assessments of the collection.

This document was prepared taking into account concepts described in the following documents:

— Data Catalog Vocabulary (DCAT)[1] [published by the World Wide Web Consortium (W3C)];

— The PROV Ontology (Prov-O)[2] [published by the World Wide Web Consortium (W3C)];

— The PROV Data Model (Prov-DM)[3] [published by the World Wide Web Consortium (W3C)];

— Project Open Data Metadata Schema v1.1[4] (published by the US Government).

Supplementary material is provided in the annexes as follows:

— Annex A provides a consolidated class hierarchy for the classes specified in this document;

— Annex B provides two examples of the registration of data sets using the facilities specified in this document;

— Annex C provides a complete description of how the concept of provenance can be captured using the facilities specified in this document.

In Clauses 6 and 7 and Annex C, this document uses:

— **bold** font to highlight terms which represent metadata objects specified by the metamodel;

— normal font for terms which represent concepts defined in Clause 3.

EXAMPLE   **Data_Set** (7.2.2.3) is a class each instance of which models a data set.

Information technology — Metadata registries (MDR) —

Part 33:
Metamodel for data set registration

1 Scope

This document provides a specification for an extension to a Metadata Registry (MDR), as specified in ISO/IEC 11179-3 in which metadata which describes data sets, collections of data available for access or download in one or more formats, can be registered. Since a set can contain a single element, this document enables the recording of metadata about a single data value.

The registered metadata provides information about the data set that can include the provenance of the data set and quality, fitness for role, and risk assessments of the data set.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 11179-3:2023, *Information technology — Metadata registries (MDR) — Part 3: Metamodel for registry common facilities*

ISO/IEC 11179-31:2023, *Information technology — Metadata registries (MDR) — Part 31: Metamodel for data specification registration*

ISO/IEC 11179-35, *Information technology — Metadata registries (MDR) — Part 35: Metamodel for model registration*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 11179-3 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
concept
unit of knowledge created by a unique combination of characteristics

Note 1 to entry: Concepts are not necessarily bound to particular natural languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.

Note 2 to entry: A concept is independent of its representation.

[SOURCE: ISO/IEC 11179-3:2023, 3.2.7]

3.2
context
circumstance, purpose and perspective under which an object is defined or used

Note 1 to entry: In this document contexts are used to represent particular programmes, projects or business areas that use, or assess, *data sets* (3.5) or *data set collections* (3.6).

[SOURCE: ISO/IEC 11179-1:2023, 3.3.3, modified — Note to entry has been added.]

3.3
access level
level of authority required from a resource owner to access a protected resource

Note 1 to entry: In the context of this document, items to which an access level may be specified are limited to a *data set* (3.5), a *data set collection* (3.6) and a *data set distribution* (3.7).

Note 2 to entry: For the public, the level of authority might describe the degree of public availability of a dataset.

EXAMPLE   Public, restricted public and non-public.

[SOURCE: ISO/IEC 2382:2015, 2126297, modified — 'an entity' has been replaced by 'a resource owner' and notes have been replaced.]

3.4
accrual periodicity
frequency with which items are added to a collection

Note 1 to entry: In the context of this document, it is assumed that an updated version of a data set is issued or made available whenever new items are added to that data set.

EXAMPLE   Annual, Bimonthly, Semiweekly, Daily, Biweekly, Semiannual, Biennial, Triennial, Three Times a Week, Three Times a Month, Continuously Updated, Monthly, Quarterly, Semimonthly, Three Times a Year, Weekly, Completely Irregular

[SOURCE: ISO 2146:2010, 7.2.6, modified — Note 1 to entry has been added.]

3.5
data set
dataset
identifiable collection of data available for access or download in one or more formats

Note 1 to entry: A data set can be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically within a larger data set. Theoretically, a data set can be as small as a single feature or feature attribute contained within a larger data set.

Note 2 to entry: A data set can be presented in a tabular form and stored and distributed in tables in word processed documents, spread sheets or databases. It could also be presented in any one of a number of alternative formats, including AVRO,[5] JSON,[6] RDF[7] and XML[8].

3.6
data set collection
curated collection of one or more *data sets* (3.5)

3.7
data set distribution
specific available form of a *data set* (3.5) or *data set collection* (3.6)

Note 1 to entry: Each data set might be available in different forms and each of these forms represents a different format of the data set or a different endpoint.

Note 2 to entry: Examples of distributions include a downloadable CSV file, an API or an RSS feed. This represents a general availability of a data set.

3.8
data set specification
formal information to describe a *data set* (3.5)

3.9
information model
graphical and textual representation of entities and the relationships between them

Note 1 to entry: An information model can exist as, at the conceptual or logical level, an entity relationship model or an object class diagram, and, at the physical level, a database schema definition.

[SOURCE: ISO/IEC 19763-12:2015, 4.2.24 modified — Note 1 to entry has been amended.]

3.10
jurisdictional coverage
jurisdiction, recognized in law as a distinct legal framework, regulatory framework or both, which is a source of rules which apply to a *data set* (3.5) or *data set collection* (3.6)

Note 1 to entry: The pivot jurisdictional coverages are United Nations (UN) recognized member states. From a legal and sovereignty perspective they are considered "peer" entities. Each UN member state (a.k.a. country) can have sub-administrative divisions as recognized jurisdictions (e.g. provinces, territories, cantons, länder, etc.) as decided by that UN member state.

Note 2 to entry: Jurisdictional coverages can be combined to form new jurisdictional coverages, (e.g. through bilateral, multilateral or international treaties). Examples are the European Union (EU), NAFTA, WTO, WCO, ICAO, WHO, Red Cross, the ISO, the IEC, the ITU, etc.

Note 3 to entry: Several levels and categories of jurisdictional coverages can exist within a jurisdictional coverage.

Note 4 to entry: The concept of "jurisdictional coverage" and its definition is derived from "jurisdictional domain" from ISO/IEC 15944-5:2008, 3.67 and is harmonized with it.

3.11
provenance
information on the place and time of origin, derivation or generation of a *data set* (3.5), proof of authenticity of the *data set*, or a record of past and present ownership of the *data set*

3.12
rights
information regarding access or restrictions based on privacy, security, intellectual property rights (IPR) or other policies

Note 1 to entry: This information may explain why a "non-public" or "restricted public" *data set* (3.5), *data set collection* (3.6) or *data set distribution* (3.7) is not "public".

Note 2 to entry: This definition is derived from Project Open Data Metadata Schema v1.1 (https://project-open -data.cio.gov/v1.1/schema/#rights).

3.13
spatial coverage
geographical area which is the subject of a *data set* (3.5) or *data set collection* (3.5)

3.14
temporal coverage
period for which a *data set* (3.5) is applicable

4 Abbreviated terms

API	Application Programming Interface
CSV	Comma-separated Values

JSON	Java Script Object Notation[6]
RDF	Resource Description Framework[7]
RSS	RDF Site Summary
URL	Uniform Resource Locator
XML	eXtensible Markup Language[8]

5 Conformance

5.1 Overview of conformance

Conformance rules for a Metadata Registry are specified in ISO/IEC 11179-3:2023, Clause 4. The subclause "Degree of Conformance" is repeated here for convenience. The subsequent subclauses extend the rules from ISO/IEC 11179-3.

5.2 Degree of conformance

5.2.1 General

The distinction between "strictly conforming" and "conforming" implementations is necessary to address the simultaneous needs for interoperability and extensions. This document describes specifications that promote interoperability. Extensions are motivated by needs of users, vendors, institutions, and industries, and:

a) are not directly specified by this document;

b) are specified and agreed to outside this document; and

c) may serve as trial usage for future editions of this document.

A strictly conforming implementation can be limited in usefulness but is maximally interoperable with respect to this document. A conforming implementation can be more useful but might be less interoperable with respect to this document.

5.2.2 Strictly conforming implementations

A strictly conforming implementation:

a) shall support all mandatory, optional and conditional classes, attributes, datatypes and associations;

b) shall not use, test, access, or probe for any extension features nor extensions to classes, attributes, datatypes, associations or any combination thereof;

c) shall not recognize, nor act on, nor allow the production of classes, attributes, datatypes, associations or any combination thereof that are dependent on any unspecified, undefined, or implementation-defined behaviour.

NOTE   The use of extensions to the metamodel can cause undefined behaviour.

5.2.3 Conforming implementations

A conforming implementation:

a) shall support all mandatory, optional and conditional classes, attributes, datatypes and associations;

b) as permitted by the implementation, may use, test, access, or probe for extension features or extensions to classes, attributes, datatypes, associations or any combination thereof;

c) may recognize, act on, or allow the production of classes, attributes, datatypes, associations or any combination thereof that are dependent on implementation-defined behaviour.

NOTE 1   All strictly conforming implementations are also conforming implementations.

NOTE 2   The use of extensions to the metamodel can cause undefined behaviour.

5.3   Conformance by feature

Conformance claims may be made to Clause 7, or to specific features within this clause. This clause is also dependent upon one or more clauses of ISO/IEC 11179-3, ISO/IEC 11179-31 and ISO/IEC 11179-35, so conformance to all or part of this clause shall be understood to imply conformance also to relevant provisions specified in one or more of the clauses in ISO/IEC 11179-3, ISO/IEC 11179-31 and ISO/IEC 11179-35.

A conformance statement shall specify exactly the features supported and not supported.

5.4   Registry conformance

5.4.1   Standard registry profiles

This document specifies the following standard profiles in addition to those specified in ISO/IEC 11179-3:2023, 4.4.2:

— **Data Set Registry**: Implements Clause 7, in addition to all provisions of the Basic registry profile of ISO/IEC 11179-3:2023, 4.4.2;

— **Data Set Registry with mapping**: Implements Clause 7, in addition to all provisions of the Basic registry with mapping profile of ISO/IEC 11179-3:2023, 4.4.2.

5.4.2   Conformance labels

Conformance to the profiles specified in 5.4.1 may be claimed using the following labels, respectively:

— ISO/IEC 11179-33:2023 Data Set Registry;

— ISO/IEC 11179-33:2023 Data Set Registry with mapping.

5.5   Implementation conformance statement (ICS)

An implementation claiming conformance to this document shall include an Implementation Conformance Statement stating:

a) whether it conforms or strictly conforms;

b) which clauses are or are not supported;

c) what extensions, if any, are supported or used.

A standard profile may be referenced, if applicable.

EXAMPLE   Product Z strictly conforms to ISO/IEC 11179-33:2023 Data Set Registry with Mapping.