INTERNATIONAL STANDARD

ISO/IEC 11179-7

First edition
2019-12

Information technology — Metadata registries (MDR) —

Part 7: Metamodel for data set registration

Reference number
ISO/IEC 11179-7:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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).

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 http://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.

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

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

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.

Introduction

Many data sets are now generally available. These may be part of a government-led open data initiative, or may 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.

In ISO/IEC 11179-3:2013, Clauses 5 to 11 the structure of a *Metadata Registry* is specified in the form of a conceptual data model. The *Metadata Registry* is used to keep information about data elements and associated concepts, such as "data element concepts", "conceptual domains" and "value domains". Generically, these are all referred to as "metadata items". Such metadata are necessary to clearly describe, record, analyse, classify and administer data. This document provides a specification of the extensions to the registry metamodel specified in ISO/IEC 11179-3:2013, Clauses 5 to 11 to enable the registration of metadata about datasets.

When considering data and metadata, it is important to distinguish between types of data/metadata, and instances of these types. ISO/IEC 11179-3:2013, Clauses 5 to 11 specifies the types of metadata objects that form the structure of a basic *Metadata Registry*. Clause 5 specifies the types of metadata objects that form an extension to that structure so that the whole structure provides facilities to keep information about data sets. A *Metadata Registry* will be populated with instances of these metadata objects (metadata items), which in turn define, for example, types of data in an application database, or, in the case of this document, data sets that are held elsewhere. Such data sets could, for example, be made available over the internet or be included as a table within a word processing document.

The facilities described in this document, together with those described in ISO/IEC 11179-3:2013, Clauses 5 to 11 provides the ability to record the following data set metadata:

— 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 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 and/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 Annexes as follows:

— Annex A provides an alphabetical list of the terms used in this document;

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

— Annex C summarizes conformance profiles specified in this document that are additional to those specified in ISO/IEC 11179-3:2013, Clause 4;

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

Information technology — Metadata registries (MDR) —

Part 7:
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 11179-3:2013, Clauses 5 to 11 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 includes the provenance and the quality of the dataset.

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 11179-3:2013, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*

ISO/IEC 11179-6, *Information technology — Metadata registries (MDR) — Part 6: Registration*

3 Terms, definitions and abbreviated terms

3.1 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 http://www.electropedia.org/

3.1.1
access level
level of authority required from an entity 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, a data set distribution and a data set collection.

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.

3.1.2
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.

[SOURCE: ISO 2146:2010, 7.2.6, modified – Note 1 to entry added and EXAMPLE deleted]

3.1.3
data element collection
collection of data elements that may be unordered or ordered

Note 1 to entry: Examples of unordered collections are a set or a bag (or multiset). An example of an ordered collection is a list.

3.1.4
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 may 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, JSON, RDF and XML.

3.1.5
data set collection
curated collection of one or more *data sets* (3.1.4)

3.1.6
data set distribution
specific available form of a *data set* (3.1.4) or *data set collection* (3.1.5)

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. It implies no information about the actual access method of the data, i.e. whether it is a direct download, API, or through accessing a Web page.

3.1.7
data set specification
additional formal information to describe a data set

Note 1 to entry: A data set specification may consist of zero, one or more ISO/IEC 11179-3 *Reference_Documents*, zero, one or more ISO/IEC 19763-12 *Information_Models* and/or zero, one or more sets of ISO/IEC 11179-3 *Data_Elements*.

3.1.8
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 amended]

3.1.9
ontology
specification of concrete or abstract things, and the relationships among them, in a prescribed domain of knowledge

Note 1 to entry: The specification should be computer processable, such as an ontology registered as metadata using the ISO/IEC 11179-3 *Concept Region* specification.

[SOURCE: ISO/IEC 19763-3:2010, 3.1.1.1, modified – Note 1 to entry amended]

3.1.10
provenance
information on the place and time of origin, derivation or generation of a resource or a record or proof of authenticity or of past ownership

3.1.11
rights
information regarding access or restrictions based on privacy, security or other policies

Note 1 to entry: This information may explain why a "non-public" or "restricted public" data set, data set distribution or data set collection 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/#dataQuality)

3.1.12
spatial coverage
geographical area which is the subject of a *data set* (3.1.4) or *data set collection* (3.1.5)

3.1.13
temporal coverage
period for which a *data set* (3.1.4) is applicable

3.2 Abbreviated terms

URL uniform resource locator

4 Conformance

4.1 Overview of conformance

Conformance rules for a Metadata Registry are specified in ISO/IEC 11179-3:2013, Clause 4. The following subclauses extend these rules.

4.2 Conformance by clause

Conformance claims may also be limited to Clause 5. This clause is also dependent upon one or more of ISO/IEC 11179-3:2013, Clauses 5 to 11, so conformance to this clause shall be understood to imply conformance also to relevant provisions specified in one or more of ISO/IEC 11179-3:2013, Clauses 5 to 11.

Conformance may, therefore, be claimed for a set of data structures and/or datatypes for Clause 5.

4.3 Standard profiles

4.3.1 General

This document specifies the following standard profiles in addition to those specified in ISO/IEC 11179-3:2013, 4.4.2.

4.3.2 Data Set Registry profile

This profile implements ISO/IEC 11179-3:2013, Clauses 7, 8 and 9 and Clause 5.

4.3.3 Extended Data Set Registry profile

This profile implements ISO/IEC 11179-3:2013, Clause 11, and also satisfies all provisions of the Data Set Registry profile.

5 Data Set package

5.1 Data Set metamodel region

5.1.1 Overview of the Data Set metamodel region

Figure 1 shows the metamodel for the registration of metadata about data sets.

Instances of each of the classes shown in Figure 1 shall be extended by one or more of an *identified item* (with its subtypes of *registered item, administered item* or *attached item*), a *designatable item* or a *classifiable item,* as specified in ISO/IEC 11179-3:2013, 5.5. See subclause 6.1.2 for details of the possible types for instances of each of the classes shown in Figure 1.

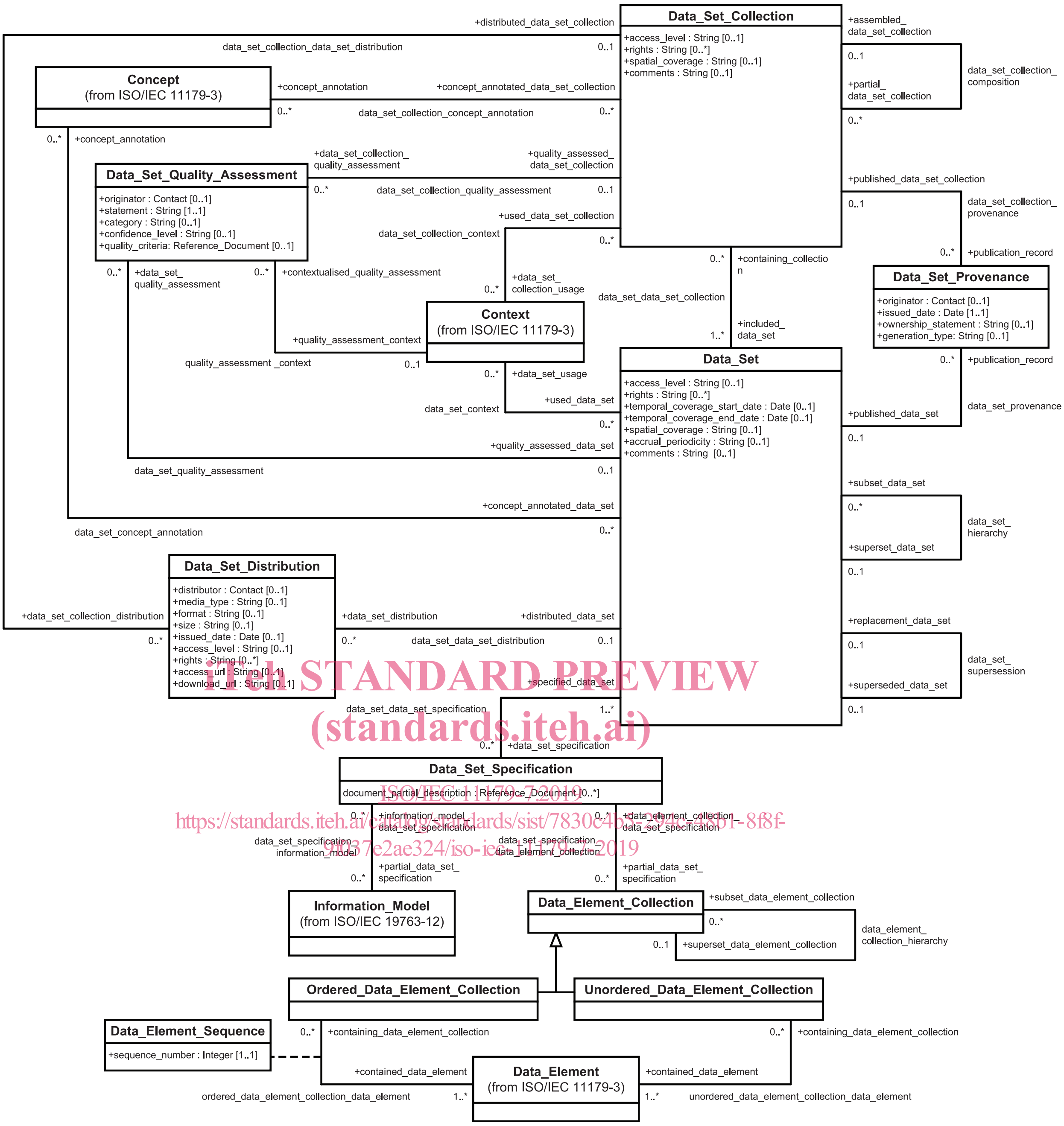

Figure 1 — Data Set metamodel region

5.1.2 Classes in the Data Set metamodel region

5.1.2.1 Concept class

The *Concept* class is described in ISO/IEC 11179-3:2013, 9.1.2.1.

In this document, the instances of the *Concept* class are used to model constructs within a concept system, that may or may not be an **ontology** (3.1.9), that are used to annotate a data set or a data set collection so as to describe the theme or category of the data set or data set collection.

A *Concept* may participate in the following additional associations specified in this document:

— *data_set_collection_concept_annotation* (5.1.4.3) with zero, one or more *concept_annotated_data_set_collection Data_Set_Collections* (5.1.2.4) in which this *Concept* is a *concept_annotation*.

— *data_set_concept_annotation* (5.1.4.8) with zero, one or more *concept_annotated_data_set Data_Sets* (5.1.2.5) in which this *Concept* is a *concept_annotation*.

The *Concept* class has no additional attributes specified in this document.

5.1.2.2 Context class

The *Context* class is described in ISO/IEC 11179-3:2013, 7.3.2.5.

In this document, the instances of the *Context* class are used to model a particular programme, project or business area that uses a data set (or collection of data sets), or a project or business area that is used as the context for a quality assessment of a data set (or collection of data sets),

A *Context* may participate in the following additional associations specified in this document:

— *data_set_context* (5.1.4.9) with zero, one or more *used_data_set Data_Sets* (5.1.2.5) in which this *Context* is a *data_set_usage*.

— *data_set_collection_context* (5.1.4.4) with zero, one or more *used_data_set_collection Data_Set_Collections* (5.1.2.6) in which this *Context* is a *data_set_collection_usage*.

— *quality_assessment_context* (5.1.4.20) with zero, one or more *contextualised_quality_assessment Data_Set_Quality_Assessments* (5.1.2.9) in which this *Context* is a *quality_assessment_context*.

The *Context* class has no additional attributes specified in this document.

5.1.2.3 Data_Element class

The *Data_Element* class is described in ISO/IEC 11179-3:2013, 11.1.2.4.

In this document, the instances of the *Data_Element* class are used to model any registered data element that is referenced in a data set specification.

A *Data_Element* may participate in the following additional associations specified in this document:

— *unordered_data_element_collection_data_element* (5.1.4.21) with zero, one or more *containing_data_element_collection Unordered_Data_Element_Collections* (5.1.2.13) in which this *Data_Element is a contained_data_element*.

— *ordered_data_element_collection_data_element* (5.1.4.19) with zero, one or more *containing_data_element_collection Ordered_Data_Element_Collections* (5.1.2.12) in which this *Data_Element* is a *contained_data_element*.

The *Data_Element* class has no additional attributes specified in this document.

5.1.2.4 Data_Element_Collection class

Data_Element_Collection is a class each instance of which models a particular **data element collection** (3.1.3), which is either an unordered collection of data elements or an ordered collection of data elements.

A data element collection brings together the data elements that are referenced in the specification of a particular data set.

A *Data_Element_Collection* shall be either an *Unordered_Data_Element_Collection* (5.1.2.13) or an *Orderered_Data_Element_Collection* (5.1.2.12) but not both.

A *Data_Element_Collection* may participate in the following associations:

— *data_element_collection_hierarchy* (5.1.4.1) with:

  — zero or one *superset_data_element_collection Data_Element_Collections* in which this *Data_Element_Collection* is a *subset_data_element_collection*.

  — zero, one or more *subset_data_element_collection Data_Element_Collections* in which this *Data_Element_Collection* is the *superset_data_element_collection*.

— *data_set_specification_data_element_collection* (5.1.4.14) with zero, one or more *data_element_collection_data_set_specification Data_Set_Specifications* (5.1.2.10) in which this *Data_Element_Collecton is a partial_data_set_specification*.

The *Data_Element_Collection* class has no attributes.

5.1.2.5 Data_Set class

5.1.2.5.1 Description of Data_Set

Data_Set is a class each instance of which models a particular **data set** (3.1.4), which is a collection of data available for access or download in one or more formats.

A *Data_Set* may participate in the following associations:

— *data_set_context* (5.1.4.9) with zero, one or more *data_set_usage Contexts* (5.1.2.2) in which this *Data_Set* is a *used_data_set*.

— *data_set_data_set_distribution* (5.1.4.11) with zero, one or more *data_set_distribution Data_Set_Distributions* (5.1.2.7) in which this *Data_Set* is the *distributed_data_set*.

— *data_set_data_set_collection* (5.1.4.10) with zero, one or more *containing_collection Data_Set_Collections* (5.1.2.6) in which this *Data_Set* is an *included_data_set*.

— *data_set_data_set_specification* (5.1.4.12) with zero, one or more *data_set_specification Data_Set_Specifications* (5.1.2.10) in which this *Data_Set* is a *specified_data_set*.

— *data_set_hierarchy* (5.1.4.13) with:

  — zero, one or more *subset_data_set Data_Sets* in which this *Data_Set is the superset_data_set*.

  — zero or one *superset_data_set Data_Set* in which this Data_*Set is a subset_data_set*.

— *data_set_concept_annotation* (5.1.4.8) with zero, one or more *concept_annotation Concepts* (5.1.2.1) in which this *Data_Set is a concept_annotated_data_set.*

— *data_set_provenance* (5.1.4.14) with zero, one or more *publication_record Data_Set_Provenances* (5.1.2.8) in which this *Data_Set is the published_data_set*.

— *data_set_quality_assessment* (5.1.4.15) with zero, one or more *data_set_quality_assessment Data_Set_Quality_Assessments* (5.1.2.9) in which this *Data_Set* is the *quality_assessed_data_set*.

— *data_set_supersession* (5.1.4.16) with:

  — zero or one *replacement_data_set Data_Set* in which this *Data_Set* is the *superseded_data_set*.

  — zero or one *superseded_data_set Data_Set* in which this *Data_Set* is the *replacement_data_set*.

The attributes of the *Data_Set* class are summarized here and specified more formally in 5.1.2.5.2:

— A *Data_Set* class may have zero or one *access_level* (5.1.2.5.2.1) of type String, which is the degree to which this data set could have been made publicly-available, regardless of whether it has been made available. If there is an instance of *Data_Set_Collection* associated with this instance of *Data_Set* and