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

 COPYRIGHT PROTECTED DOCUMENT

Contents

Page

List of Figures

List of Tables

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.

5.6 Obligation

Properties and relationships specified in this document are one of: Mandatory, Conditional or Optional. The obligation is not explicitly stated but is to be inferred from the multiplicity of the property or relationship, and the presence or absence of a condition.

For the purpose of conformance:

a) mandatory properties and relationships shall exist and shall conform to the provisions of this document;

b) anything specified as Conditional within this document shall be treated as Mandatory if the associated condition is satisfied and shall otherwise be not present;

c) optional properties and relationships are not required to exist, but if they do exist they shall conform to the provisions of this document.

Such obligation is enforced if and only if the Registration Status of the associated registry items is Recorded or higher (see ISO/IEC 11179-3:2023, 3.2.92 and 9.4.6.3 and ISO/IEC 11179-6:2023, 4.3.2[9]).

6 Relationship to ISO/IEC 11179-3:2023

6.1 Metamodel for a metadata registry

A metamodel is a model that describes other models. A metamodel provides a mechanism for understanding the precise structure and components of the specified models, which are needed for the successful sharing of the models by users, software facilities or some combination.

ISO/IEC 11179-3 uses a metamodel to describe the information model of a metadata registry. The registry in turn will be used to describe and model other data, for example about enterprise, public administration or business applications. The registry metamodel is specified as a conceptual data model, i.e. one that describes how relevant information is structured in the natural world. In other words, it is how the human mind is accustomed to thinking of the information.

6.2 Specification of the metamodel

The conventions used in specifying the metamodel are described in ISO/IEC 11179-3:2023, 5.3. All the classes specified in the Data_Set package (see Clause 7) inherit from **Item**, which is specified in the Core model region in ISO/IEC 11179-3:2023, 6.4.2.1. As **Items**, instances of the classes specified in the Data_Set package may be identified, registered, administered, named, defined and classified.

6.3 Use of UML Class diagrams and textual description

This document uses both text and UML class diagrams (see References [10] and [11]) to describe the metamodel. Both the text and UML class diagrams are normative and are intended to be complementary. However, if a conflict exists between what is specified in the UML Class diagrams and what is specified in text, the text takes precedence until a correction is made to make them consistent. Further, if a conflict exists between a formal definition and other normative text, the formal definition takes precedence until a correction is made to make them consistent.

A consolidated UML class hierarchy is included as Annex A.

While the model diagrams are presented in UML class diagram notation, this document does not assume nor endorse any specific system environment, database management system, database design paradigm, system development methodology, data definition language, command language, system interface, user interface, computing platform, or any technology required for implementation.

In this clause, Clause 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.

6.4   Package dependencies

Figure 1 — Package dependencies

Figure 1 illustrates the dependencies among the packages. The lines in the figure illustrate dependencies in the direction of the arrow. The Data_Set package is specified in Clause 7. The other packages shown in Figure 1 are specified in ISO/IEC 11179-3, ISO/IEC 11179-31, or ISO/IEC 11179-35. In order to implement a package that has dependencies, the packages on which it is dependent shall also be implemented. The dependencies are of three types:

a) subclassing from classes in another package, e.g. **Data_Set** (7.2.2.3) and **Data_Set_Distribution** (7.2.2.5) in the Data_Set metamodel package (7) are subclassed from the **Item** class in the Basic_and_Core metamodel package in ISO/IEC 11179-3;

b) relationship between classes, e.g. **Data_Set** (7.2.2.3) in the Data_Set metamodel package (7) has a relationship through the **data_set_concept_annotation** (7.2.3.1) association with **Concept** in the Basic_and_Core metamodel package in ISO/IEC 11179-3;

c) some attributes use a predefined datatype or a class from another package as a datatype, e.g. the **originator** attribute of the **Data_Set_Assessment** (7.2.2.7) class in the Data_Set metamodel package uses the **Contact** class of the Basic_and_Core metamodel package in ISO/IEC 11179-3 as a data type.

Conformance options are specified in Clause 5 and standard conformance profiles in 5.4.

7 Data_Set package

7.1 Overview of the Data_Set package

The Data_Set package consists of a single metamodel region, the Data_Set metamodel region.

7.2 Data_Set metamodel region

7.2.1 Overview of the Data_Set metamodel region

Figure 2 shows the metamodel for the registration of metadata about data sets.

Figure 2 — Data_Set metamodel region

Examples of the registration of data sets using this metamodel are included at Annex B.

7.2.2 Classes in the Data_Set metamodel region

7.2.2.1 Concept class

7.2.2.1.1 Direct superclass

Concept is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.1.2   Description of Concept

Concept is part of the Core metamodel region and is specified in ISO/IEC 11179-3:2023, 6.4.2.2. Additional associations are specified in this metamodel region.

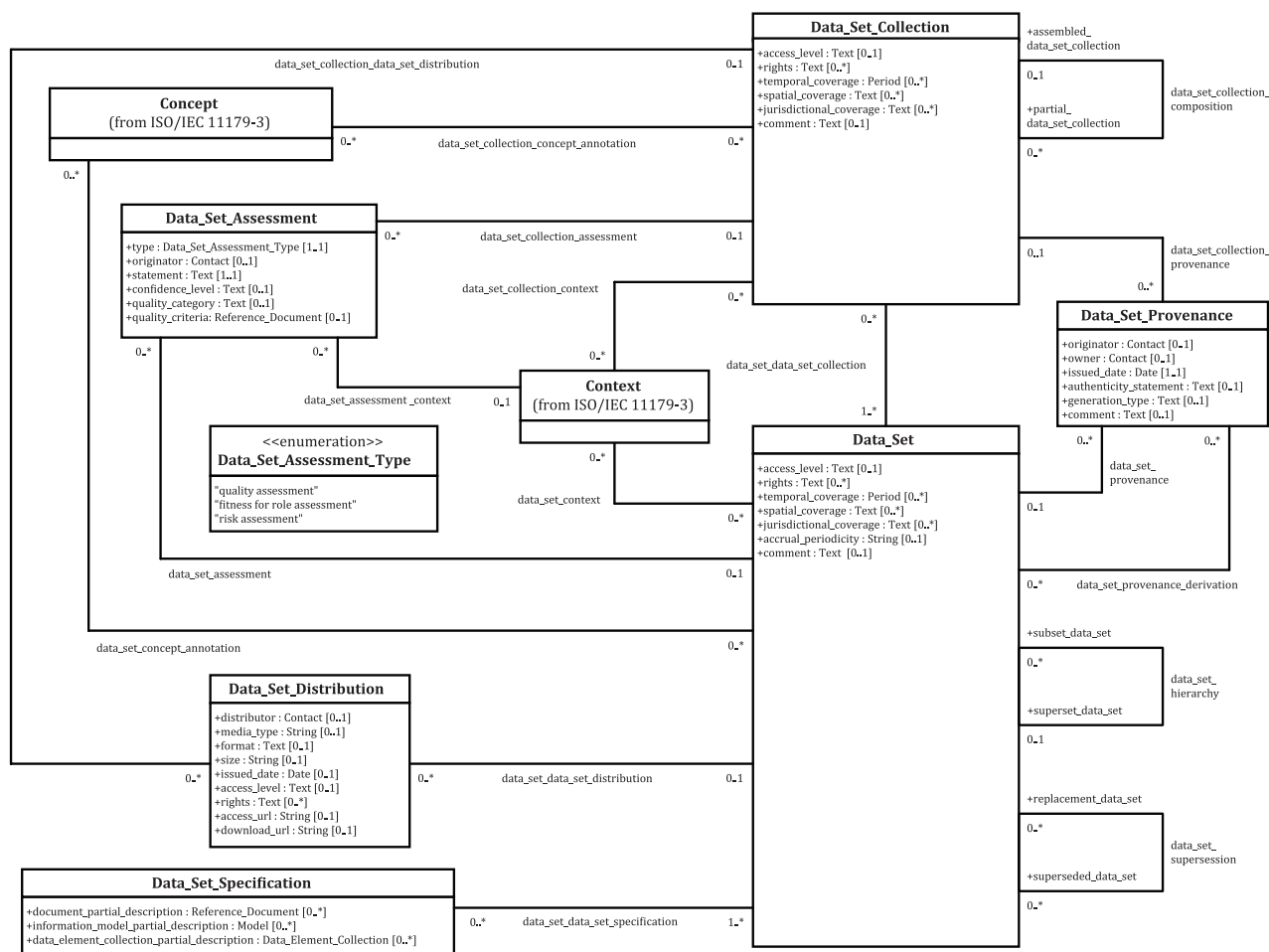**Concept** is a class each instance of which models a concept, a unit of knowledge created by a unique combination of characteristics. A concept is independent of representation.

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, that are used to annotate a data set or a data set collection to describe the theme or category of the data set or data set collection.

7.2.2.1.3   Associations of Concept

As a subclass of **Item**, **Concept** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_concept_annotation** (7.2.3.1);

— **data_set_collection_concept_annotation** (7.2.3.2).

7.2.2.1.4   Attributes of Concept

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

7.2.2.2   Context class

7.2.2.2.1   Direct superclass

Context is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.2.2   Description of Context

Context is part of the Core metamodel region and is specified in ISO/IEC 11179-3:2023, 6.4.2.3. Additional associations are specified in this metamodel region.

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 data set collection, or a programme, project or business area that is used as the context for an assessment of a data set (or data set collection).

7.2.2.2.3   Associations of Context

As a subclass of **Item**, **Context** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_context** (7.2.3.3);

— **data_set_collection_context** (7.2.3.4);

— **data_set_assessment_context** (7.2.3.5).

7.2.2.2.4   Attributes of Context

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

7.2.2.3   Data_Set class

7.2.2.3.1   Direct superclass

Data_Set is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.3.2   Description of Data_Set

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

7.2.2.3.3   Associations of Data_Set

As a subclass of **Item**, **Data_Set** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_concept_annotation** (7.2.3.1);

— **data_set_context** (7.2.3.3);

— **data_set_data_set_collection** (7.2.3.6);

— **data_set_data_set_distribution** (7.2.3.7);

— **data_set_data_set_specification** (7.2.3.8);

— **data_set_hierarchy** (7.2.3.9);

— **data_set_provenance** (7.2.3.10);

— **data_set_provenance_derivation** (7.2.3.11);

— **data_set_assessment** (7.2.3.12);

— **data_set_supersession** (7.2.3.13).

7.2.2.3.4   Attributes of Data_Set

The attributes of the **Data_Set** class are specified in Table 1.

Table 1 — Attributes of the Data_Set class

Attribute name	Multiplicity	Datatype	Description
access_level	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: 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** (7.2.2.4) linked to this instance of **Data_Set** and the **access_level** of the linked instance of **Data_Set_Collection** differs from the **access_level** of this instance of **Data_Set** then it is assumed that the **access_level** of this instance of **Data_Set** overrides the **access_level** of the linked instance of **Data_Set_Collection**. Examples: "public", "restricted-public", "non-public".

Table 1 *(continued)*

Attribute name	Multiplicity	Datatype	Description
rights	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: Information regarding any access or restrictions based on privacy, security, intellectual property rights (IPR) or other policies that apply to this data set. This should also serve as an explanation for the selected **access_level**. If there is an instance of **Data_Set_Collection** ([7.2.2.4](#)) linked to this instance of **Data_Set** and the **rights** of the linked instance of **Data_Set_Collection** differs from the **rights** of this instance of **Data_Set** then it is assumed that the **rights** of this instance of **Data_Set** overrides the **rights** of the linked instance of **Data_Set_Collection**.
temporal_coverage	0..*	**Datetime_Period** (ISO/IEC 11179-3:2023, 6.3.10)	Definition: The period or periods for which this data set is applicable. If there is an instance of **Data_Set_Collection** ([7.2.2.4](#)) linked to this instance of **Data_Set** and the **temporal_coverage** *of* the linked instance of **Data_Set_Collection** differs from the **temporal_coverage** of this instance of **Data_Set** then it is assumed that the **temporal_coverage** of this instance of **Data_Set** overrides the **temporal_coverage** of the linked instance of **Data_Set_Collection**.
spatial_coverage	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The spatial applicability or spatial applicabilities of this data set. Each spatial applicability is a geographical area covered by the data set. This could include a named place or a description of the boundary of a spatial region. If there is an instance of **Data_Set_Collection** ([7.2.2.4](#)) linked to this instance of **Data_Set** and the **spatial_coverage** *of* the linked instance of **Data_Set_Collection** differs from the **spatial_coverage** of this instance of **Data_Set** then it is assumed that the **spatial_coverage** of this instance of **Data_Set** overrides the **spatial_coverage** of the linked instance of **Data_Set_Collection**.
jurisdictional_coverage	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The jurisdiction or jurisdictions, each recognized in law as a distinct legal framework, regulatory framework or both, which is a source of rules which apply to this data set. If there is an instance of **Data_Set_Collection** ([7.2.2.4](#)) linked to this instance of **Data_Set** and the **jurisdictional_coverage** of the linked instance of **Data_Set_Collection** differs from the **jurisdictional_coverage** of this instance of **Data_Set** then it is assumed that the **jurisdictional_coverage** of this instance of **Data_Set** overrides the **jurisdictional_coverage** of the linked instance of **Data_Set_Collection**.
accrual_periodicity	0..1	**String** (ISO/IEC 11179-3:2023, 6.2.11)	Definition: The frequency at which new, revised or updated versions of this data set are made available.
comment	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: Additional information that is known about the data set.

7.2.2.3.5 Constraints on Data_Set

— Constraint 1: Each instance of the **Data_Set** class shall exist only if at least one instance of the **item_designation** association of the parent **Item** class shall be instantiated.

— Constraint 2: If there is no linked instance of the **Data_Set_Specification** class (7.2.2.8) through the **data_set_data_set_specification** association (7.2.3.8) for an instance of the **Data_Set** class, at least one instance of the **item_definition** association of the parent instance of the **Item** class shall be instantiated.

— Constraint 3: If there is no definition supplied by the existence of an instance of the **item_definition** association of the parent instance of the **Item** class for an instance of the **Data_Set** class, at least one instance of the **Data_Set_Specification** class (7.2.2.8) shall be provided through the instantiation of a linked instance of the **data_set_data_set_specification** association (7.2.3.8).

7.2.2.4 Data_Set_Collection class

7.2.2.4.1 Direct superclass

Data_Set_Collection is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.4.2 Description of Data_Set_Collection

Data_Set_Collection is a class each instance of which models a particular data set collection, which is a curated collection of one or more data sets. Such a collection can take the form of a catalogue of data sets.

7.2.2.4.3 Associations of Data_Set_Collection

As a subclass of **Item**, **Data_Set_Collection** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_collection_concept_annotation** (7.2.3.2);

— **data_set_collection_context** (7.2.3.4);

— **data_set_data_set_collection** (7.2.3.6);

— **data_set_collection_composition** (7.2.3.14);

— **data_set_collection_data_set_distribution** (7.2.3.15);

— **data_set_collection_provenance** (7.2.3.16);

— **data_set_collection_assessment** (7.2.3.17).

7.2.2.4.4 Attributes of Data_Set_Collection

The attributes of the **Data_Set_Collection** class are specified in Table 2.

Table 2 — Attributes of the Data_Set_Collection class

Attribute name	Multiplicity	Datatype	Description
access_level	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The degree to which this data set collection could have been made publicly-available, regardless of whether it has been made available. Examples: "public", "restricted-public", "non-public".

Table 2 *(continued)*

Attribute name	Multiplicity	Datatype	Description
rights	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: Information regarding any access or restrictions based on privacy, security, intellectual property rights (IPR) or other policies that apply to this data set collection. This should also serve as an explanation for the selected **access_level.**
temporal_ coverage	0..*	**Datetime_Period** (ISO/IEC 11179-3:2023, 6.3.10)	Definition: The period or periods for which this data set collection is applicable.
spatial_ coverage	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The spatial applicability or spatial applicabilities of this data set collection. These are the geographical areas covered by the data sets included within the data set collection. This could include a named place or a description of the boundary of a spatial region.
jurisdictional_ coverage	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The jurisdiction or jurisdictions, each recognized in law as a distinct legal, regulatory framework or both, which is a source of rules which apply to some or all of the data sets included within this data set collection.
comment	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: Additional information that is known about the data set collection.

7.2.2.4.5 Constraint on Data_Set_Collection

Each instance of the **Data_Set_Collection** class shall exist only if at least one instance of the **item_ designation** association of the parent **Item** class shall be instantiated.

7.2.2.5 Data_Set_Distribution class

7.2.2.5.1 Direct superclass

Data_Set_Distribution is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.5.2 Description of Data_Set_Distribution

Data_Set_Distribution is a class each instance of which models a particular data set distribution, which is a specific available form of a data set or a data set collection.

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

7.2.2.5.3 Associations of Data_Set_Distribution

As a subclass of **Item**, **Data_Set_Distribution** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_data_set_distribution** (7.2.3.7);

— **data_set_collection_data_set_distribution** (7.2.3.15).

7.2.2.5.4 Attributes of Data_Set_Distribution

The attributes of the **Data_Set_Distribution** class are specified in Table 3.

Table 3 — Attributes of the Data_Set_Distribution class

Attribute name	Multiplicity	Datatype	Description
distributor	0..1	**Contact** (ISO/IEC 11179-3:2023, 6.3.5)	Definition: The individual or organization responsible for the publication of the data set or data set collection that is linked to this data set distribution.
media_type	0..1	**String** (ISO/IEC 11179-3:2023, 6.2.11)	Definition: The machine-processable file format of this data set distribution as defined by the Internet Assigned Numbers Authority (IANA)[12].
format	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: A human-readable description of the file format of this data set distribution.
size	0..1	**String** (ISO/IEC 11179-3:2023, 6.2.11)	Definition: The size of this data set distribution expressed in bytes or multiples of bytes. This may be an approximation when the precise size is not known
issued_date	0..1	**Date** (ISO/IEC 11179-3:2023, 6.2.4)	Definition: The date on which this data set distribution was formally issued or made available.
access_level	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The degree to which this data set distribution could have been made publicly-available, regardless of whether it has been made available Examples: "public", "restricted-public", "non-public".
rights	0..*	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: Information regarding any access or restrictions based on privacy, security, intellectual property rights (IPR) or other policies that apply to this data set distribution. This should also serve as an explanation for the value assigned to the **access_level** attribute.
access_url	0..1	**String** (ISO/IEC 11179-3:2023, 6.2.11)	Definition: The URL for an indirect means of accessing the linked data set via this data set distribution.
download_url	0..1	**String** (ISO/IEC 11179-3:2023, 6.2.11)	Definition: The URL for a direct means of accessing the related data set via this data set distribution.

7.2.2.5.5 Constraint on Data_Set_Distribution

An instance of the **Data_Set_Distribution** class shall exist only if at least one of the **data_set_data_set_distribution** (7.2.3.7) association or the **data_set_collection_data_set_distribution** (7.2.3.15) association is instantiated.

7.2.2.6 Data_Set_Provenance class

7.2.2.6.1 Direct superclass

Data_Set_Provenance is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.6.2 Description of Data_Set_Provenance

Data_Set_Provenance is a class each instance of which models aspects of a particular provenance of a data set or data set collection, which is a record of information on the place and time of origin, derivation or generation of the data set or data set collection.

The full concept of provenance of a data set or data set collection includes proof of authenticity or of past ownership. Instances of the **Data_Set_Provenance** class do not provide all the details required to determine the full provenance of a data set or data set collection. A complete view of how the provenance concept can be applied to data sets or data set collections using this metamodel is provided in Annex C.

7.2.2.6.3 Associations of Data_Set_Provenance

As a subclass of **Item**, **Data_Set_Provenance** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_provenance** (7.2.3.10);

— **data_set_provenance_derivation** (7.2.3.11);

— **data_set_collection_provenance** (7.2.3.16).

7.2.2.6.4 Attributes of Data_Set_Provenance

The attributes of the **Data_Set_Provenance** class are specified in Table 4.

Table 4 — Attributes of the Data_Set_Provenance class

Attribute name	Multiplicity	Datatype	Description
originator	0..1	**Contact** (ISO/IEC 11179-3:2023, 6.3.5)	Definition: The individual or organization responsible for the origination of the data set or data set collection which is the subject of this provenance.
owner	0..1	**Contact** (ISO/IEC 11179-3:2023, 6.3.5)	Definition: The individual or organization that claims ownership of the data set or data set collection which is the subject of this provenance.
issued_date	1..1	**Date** (ISO/IEC 11179-3:2023, 6.2.4)	Definition: The date on which the linked version of the data set or data set collection was issued.
authenticity_ statement	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: A statement that describes how the proof of authenticity of the linked data set or data set collection was obtained. The proof of authenticity could be obtained using a digital signature or some other mechanism.
generation_ type	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: A statement that describes how the linked data set or data set collection was generated. Examples: "generated by sensors" and "derived from observations".
comment	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: Additional information that helps to explain the provenance of the associated data set or data set collection.

7.2.2.6.5 Constraint on Data_Set_Provenance

An instance of the **Data_Set_Provenance** class shall exist only if at least one of the **data_set_ provenance** (7.2.3.10) association or the **data_set_collection_provenance** (7.2.3.16) association is instantiated.

7.2.2.7 Data_Set_Assessment class

7.2.2.7.1 Direct superclass

Data_Set_Assessment is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.7.2 Description of Data_Set_Assessment

Data_Set_Assessment is a class each instance of which models a particular assessment about a data set or a data set collection.

This assessment shall be an assessment of one of the following:

— the quality of the data set or data set collection;

— the fitness for role of the data set or data set collection; or

— the risks involved in using the data set or data set collection.

If this assessment is an assessment of the quality of a data set or data set collection, it can be based on a particular quality dimension or can be a more general assessment of quality. The quality dimensions against which quality is assessed can be derived from international standards, such as ISO 8000-8[13], ISO 19157[14] and ISO/IEC 25012[15] or can be from more informal schemes. Examples of quality dimensions in more informal schemes include "accuracy", "relevance", "consistency" and "completeness".

7.2.2.7.3 Associations of Data_Set_Assessment

As a subclass of **Item**, **Data_Set_Assessment** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following associations:

— **data_set_assessment_context** (7.2.3.5);

— **data_set_assessment** (7.2.3.12);

— **data_set_collection_assessment** (7.2.3.17).

7.2.2.7.4 Attributes of Data_Set_Assessment

The attributes of the **Data_Set_Assessment** class are specified in Table 5.

Table 5 — Attributes of the Data_Set_Assessment class

Attribute name	Multiplicity	Datatype	Description
type	1..1	**Data_Set_Assessment_Type** (7.2.4)	Definition: The type of this assessment of the linked data set or data set collection. Condition: The value specified for this attribute shall be one of those specified for the **Data_Set_Assessment_Type** enumeration class (7.2.4.1).
originator	0..1	**Contact** (ISO/IEC 11179-3:2023, 6.3.5)	Definition: The individual or organization responsible for the assessment of the linked data set or data set collection.
statement	1..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: A statement of the assessment of the quality of, the fitness for role of, or the risk associated with using the data within the linked data set or the data within the data sets included in the linked data set collection.

Table 5 *(continued)*

Attribute name	Multiplicity	Datatype	Description
confidence_ level	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The degree of confidence, if any, of this assessment of the data within the linked data set or the data within the data sets included in the linked data set collection.
quality_ category	0..1	**Text** (ISO/IEC 11179-3:2023, 6.2.12)	Definition: The quality dimension against which the assessment of the quality of the data within the linked data set or the data within the data sets included in the linked data set collection. is assessed. Condition: A **quality_category** shall be specified only if the value of the **type** attribute is "quality assessment".
quality_criteria	0..1	**Reference_ Document** (ISO/IEC 11179-3:2023, 6.3.8)	Definition: The document that specifies the criteria against which this assessment of the quality of the data within the linked data set or the data within the data sets included in the linked data set collection has been assessed. Condition: A **quality_criteria** shall be specified only if the value of the **type** attribute is "quality assessment".

7.2.2.7.5 Constraints on Data_Set_Assessment

An instance of the **Data_Set_Assessment** class shall exist only if at least one of the **data_set_ assessment** (7.2.3.12) association or the **data_set_collection_assessment** (7.2.3.17) association is instantiated.

7.2.2.8 Data_Set_Specification class

7.2.2.8.1 Direct superclass

Data_Set_Specification is a subclass of **Item** (specified in ISO/IEC 11179-3:2023, 6.4.2.1), allowing instances to be identified, registered, administered, named, defined and classified.

7.2.2.8.2 Description of Data_Set_Specification

Data_Set_Specification is a class each instance of which models a particular data set specification, which provides additional formal information to describe a data set.

A data set specification may consist of zero, one or more reference documents, zero, one or more information models, each of which shall be in accordance with either ISO/IEC 11179-35 or ISO/IEC 19763-12, zero, one or more collections of data elements, each of which shall be in accordance with ISO/IEC 11179-31:2023, 7.9.2.3, or some combination.

7.2.2.8.3 Associations of Data_Set_Specification

As a subclass of **Item**, **Data_Set_Specification** inherits **Item**'s associations (ISO/IEC 11179-3:2023, 6.4.2.2.3). This metamodel region adds the following association:

— **data_set_data_set_specification** (7.2.3.8).

7.2.2.8.4 Attributes of Data_Set_Specification

The attributes of the Data_Set_Specification class are specified in Table 6.

Table 6 — Attributes of the Data_Set_Specification class

Attribute name	Multiplicity	Datatype	Description
document_partial_ description	0..*	**Reference_ Document** (ISO/IEC 11179-3:2023, 6.3.8)	Definition: A document that provides some additional description of the data set that is the subject of the data set specification.
information_model_ partial_description	0..*	**Model** (ISO/IEC 11179-35:2023, 7.2.2.2)	Definition: An information or data model that pro-vides some additional description of the data set that is the subject of the data set specification.
data_element_ collection_partial_ description	0..*	**Data_Element_ Collection** (ISO/IEC 11179-31:2023, 7.9.2.3)	Definition: A collection of data elements that provides some additional description of the data set that is the subject of the data set specification.

7.2.2.8.5   Constraint on Data_Set_Specification

An instance of the **Data_Set_Specification** class shall exist only if at least one of the following conditions applies:

— there exists at least one value for the **document_partial_description** attribute;

— there exists at least one value for the **information_model_partial_description** attribute;

— there exists at least one value for the **data_element_collection_partial_description** attribute.

7.2.2.8.6   Additions to Data_Set_Specification

Where there is a requirement to add additional domain or content specific 'attributes' to an instance of the **Data_Set_Specification** class, this can be achieved using the **Slot** class specified in ISO/IEC 11179-3:2023, 6.4.2.4.

7.2.3   Associations in the Data_Set metamodel region

7.2.3.1   data_set_concept_annotation association

The **data_set_concept_annotation** association records the binding of zero, one or more instances of the **Concept** (7.2.2.1) class to zero, one or more instances of the **Data_Set** (7.2.2.3) class.

This association registers the concepts that are used to annotate a data set to describe the theme or category of the data set.

7.2.3.2   data_set_collection_concept_annotation association

The **data_set_collection_concept_annotation** association records the binding of zero, one or more instances of the **Concept** (7.2.2.1) class to zero, one or more instances of the **Data_Set_Collection** (7.2.2.4) class.

This association registers the concepts that are used to annotate a data set collection to describe the theme or category of each of the data sets within the data set collection.

7.2.3.3   data_set_context association

The **data_set_context** association records the binding of zero, one or more instances of the **Context** (7.2.2.2) class to zero, one or more instances of the **Data_Set** (7.2.2.3) class.

This association registers the contexts that are applied to a data set, with each context representing a programme, project or business area that uses the data set.

7.2.3.4 data_set_collection_context association

The **data_set_collection_context** association records the binding of zero, one or more instances of the **Context** (7.2.2.2) class to zero, one or more instances of the **Data_Set_Collection** (7.2.2.4) class.

This association registers the contexts that are applied to a data set collection, with each context representing a programme, project or business area that uses all of the data sets within the data set collection.

7.2.3.5 data_set_assessment_context association

The **data_set_assessment_context** association records the binding of zero, one or more instances of the **Data_Set_Assessment** (7.2.2.7) class to zero or one instance of the **Context** (7.2.2.2) class.

This association registers the context, a project or business area, within which the associated assessment of a data set or data set collection was made.

7.2.3.6 data_set_data_set_collection association

The **data_set_data_set_collection** association records the binding of one or more instances of the **Data_Set** (7.2.2.3) class to zero, one or more instances of the **Data_Set_Collection** (7.2.2.4) class.

This association registers the data set collection to which a data set belongs.

7.2.3.7 data_set_data_set_distribution association

The **data_set_data_set_distibution** association records the binding of one and only one instance of the **Data_Set** (7.2.2.3) class to zero, one or more instances of the **Data_Set_Distribution** (7.2.2.4.5) class.

This association registers the data set distribution that makes the linked data set accessible.

7.2.3.8 data_set_data_set_specification association

The **data_set_data_set_specification** association records the binding of zero, one or more instances of the **Data_Set** (7.2.2.3) class to zero, one or more instances of the **Data_Set_Specification** (7.2.2.8) class.

This association registers any data set specification applied to a data set, where the data set specification provides additional formal information to describe a data set.

7.2.3.9 data_set_hierarchy association

The **data_set_hierarchy** association is a reflexive association which records the binding of zero or one instance of the **Data_Set** (7.2.2.3) class, each of which represents a superset, to zero, one or more other instances of the **Data_Set** class, each of which represents a subset.

This association registers any superset-subset hierarchy that exists within a group of data sets. This hierarchy can then be used to drill down to further detail.

7.2.3.10 data_set_provenance association

The **data_set_provenance** association records the binding of zero, one or more instances of the **Data_Set_Provenance** (7.2.2.6) class to zero or one instance of the **Data_Set** (7.2.2.3) class.

This association registers the provenance of a data set.

7.2.3.11 data_set_provenance_derivation association

The **data_set_provenance_derivation** association records the binding of zero, one or more instances of the **Data_Set_Provenance** (7.2.2.6) class to zero, one or more instances of the **Data_Set** (7.2.2.3) class.

This association registers the group of data sets from which the provenance of another data set (linked through the **data_set_provenance** association (7.2.3.10)) is derived.

7.2.3.12 data_set_assessment association

The **data_set_assessment** association records the binding of zero, one or more instances of the **Data_Set_Assessment** (7.2.2.7) class to zero or one instance of the **Data_Set** (7.2.2.3) class.

This association registers an assessment, a quality assessment, a fitness for role assessment or a risk assessment, that is applied to the linked data set.

7.2.3.13 data_set_supersession association

The **data_set_supersession** association is a reflexive association which records the binding of zero, one or more instances of the **Data_Set** (7.2.2.3) class, each of which represents a superseded data set, to zero, one or more other instances of the **Data_Set** class, each of which represents a replacement data set.

This association registers that a single data set, or many data sets, (the superseded data set(s)) has been replaced by single other data set, or many other data sets, (the replacement data set(s)).

7.2.3.14 data_set_collection_composition association

The **data_set_collection_composition** association is a reflexive association which records the binding of zero, one or more instances of the **Data_Set_Collection** (7.2.2.4) class, each of which represents a partial data set collection, to zero or one other instance of the **Data_Set_Collection** class, representing the assembled data set collection.

This association registers the data set collections (the partial data set collections) that comprise another data set collection (the assembled data set collection).

7.2.3.15 data_set_collection_data_set_distribution association

The **data_set_collection_data_set_distribution** association records the binding of zero or one instance of the **Data_Set_Collection** (7.2.2.4) class to zero, one or more instances of the **Data_Set_Distribution** (7.2.2.4.5) class.

This association registers the data set distributions that apply to each of the data sets within the data set collection, thus recognising that it is through each of these data set distributions that all the data sets within the data set collection are made accessible.

7.2.3.16 data_set_collection_provenance association

The **data_set_collection_provenance** association records the binding of zero, one or more instances of the **Data_Set_Provenance** (7.2.2.6) class to zero or one instance of the **Data_Set_Collection** (7.2.2.4) class.

This association registers the provenance of a data set collection, with this provenance applying to all the data sets within the data set collection.

7.2.3.17 data_set_collection_assessment association

The **data_set_collection_assessment** association records the binding of zero, one or more instances of the **Data_Set_Assessment** (7.2.2.7) class to zero or one instances of the **Data_Set_Collection** (7.2.2.4) class.

This association registers a data set assessment, a quality assessment, a fitness for role assessment or a risk assessment, that is applied to the associated data set collection. This assessment will generally apply to each of the data sets within the data set collection. Any data set assessments linked to a data set within a data set collection through an instance of the **data_set_assessment** association (7.2.3.12) take precedence over any data set assessment linked to the data set collection.

7.2.4 Datatypes in the Data_Set metamodel region

7.2.4.1 Data_Set_Assessment_Type enumerated class

Data_Set_Assessment_Type is an enumerated class. This enumerated class is used as the datatype for the **type** attribute of the **Data_Set_Assessment** (7.2.2.7) class.

The values of the **Data_Set_Assessment_Type** enumerated class are specified in Table 7.

Table 7 — Values in the Data_Set_Assessment_Type enumerated class

Value	Description
"quality assessment"	Indicates that this assessment is an assessment of the quality of the data set or data set collection within the given context.
"fitness for role assessment"	Indicates that this assessment is an assessment of the fitness for role of the data set or data set collection within the given context.
"risk assessment"	Indicates that this assessment is an assessment of the risks involved in using the data set or data set collection within the given context.

Annex A
(informative)

Consolidated Class Hierarchy

Figure A.1 shows all classes specified in this document, or referenced from ISO/IEC 11179-3, that participate in a class hierarchy. Classes that do not participate in this class hierarchy are not shown.
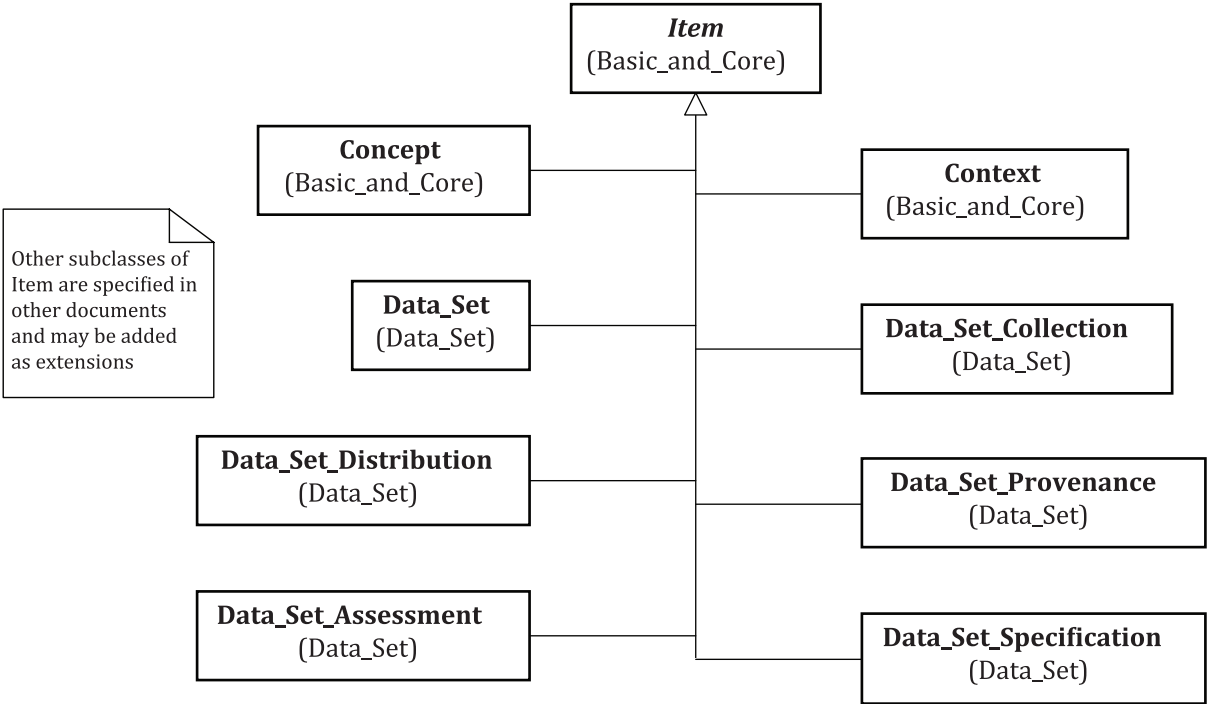

Figure A.1 — Consolidated Class Hierarchy

NOTE   Standards or implementations which extend the metamodel in this document can extend this class hierarchy.

Annex B
(informative)

Examples of data set registration

B.1 Overview

Two examples of the registration of data sets are provided in this annex. The first example is of a data set that is routinely updated every month. The second example shows the registration of a historic data set that is then decomposed to show the registration of a single value.

B.2 'Air Quality' data set example

B.2.1 Overview of the 'Air Quality' example

This example data set, of air quality data for London, UK, published by the Greater London Authority (GLA) is taken from the following website (accessed on 23 October 2016):

https://data.london.gov.uk/dataset/london-average-air-quality-levels/resource/acce7f88-70f0-4fd0-9160-f02a9d96b2c3.

The data is published in a spreadsheet workbook, comprised of five worksheets, that is updated monthly. Table B.1 shows an extract from the 'Metadata' worksheet and Table B.2 shows an extract from the 'Monthly Averages' worksheet.

Table B.1 — Extract from the 'Metadata' worksheet for the 'Air Quality' example

Name	London average air quality levels
ShortName	Air Quality London
Theme	Environment
Sub-theme	Air
Title	Average Air Quality in London
Description	Roadside and background average readings for nitrogen dioxide, nitric oxide, oxides of nitrogen, ozone, particulate matter (PM10 and PM2.5), and sulphur dioxide.
Subject	Air quality
Subject.keyword	NO2, SO2, PM10, PM2.5, O3, levels, daily, hourly, index, monitoring, NO, Nox
Publisher	GLA
Date.available	January 2008 onwards
Creator	King's College London
Date.created	Oct 16
Coverage.spatial	London
Coverage.temporal	Hourly
Type	Measurements
Language	English
Rights	Crown Copyright
More info	http://www.londonair.org.uk

Table B.1 *(continued)*

Download from	Raw data available under licence from King's College only
Measure	Micrograms per cubic metre of air (ug/m^3)
Warnings/Notes	In the Mayor's Air Quality Strategy, the Mayor adopted a policy to support the improvement of the air quality monitoring network in London so that it provides accurate and up to date information about air quality trends and current conditions in London. Publishing this data forms part of the delivery of that proposal, and also part of a larger push by the GLA towards transparency and sharing of information.
	Each pollutant is measured once per hour across the London Air Quality Network. Averages are calculated from these measurements.
	The averages shown here have been placed into DEFRA air quality index bands but are indicative only, and do not represent official air quality monitoring for European legislation.
	The revised Daily Air Quality Index, published by DEFRA in April 2013 is used within this spreadsheet.
Next release	Nov 16

Table B.2 — Extract from the 'Monthly Averages' worksheet for the 'Air Quality' example

Month	London mean roadside						
	Nitric oxide (ug/m^3)	Nitrogen dioxide (ug/m^3)	Oxides of nitrogen (ug/m^3)	Ozone (ug/m^3)	PM10 particulate (ug/m^3)	PM2.5 particulate (ug/m^3)	Sulphur dioxide (ug/m^3)
Jan-2008		55.5		29.5	25.0	14.7	4.2
Feb-2008		75.9		20.3	39.5	28.8	7.6
Mar-2008		55.6		40.1	21.6	12.3	3.9
Apr-2008		61.8		37.9	28.7	20.5	4.5
May-2008		62.9		46.3	34.6	27.5	4.6
Jun-2008		49.2		39.8	23.2	16.0	3.6
Jul-2008		48.4		35.0	23.0	14.2	3.1
Aug-2015	69.7	49.0	118.8	29.0	20.5	12.2	0.5
Sep-2015	99.3	57.0	156.2	21.2	20.4	12.6	1.0
Oct-2015	108.8	58.5	167.3	18.3	26.0	18.6	0.0
Nov-2015	98.3	53.9	152.2	24.9	21.8	12.7	3.0
Dec-2015	72.4	47.2	119.7	32.6	24.1	11.9	1.6
Jan-2016	123.1	60.0	183.2	23.8	23.2	13.6	3.9
Feb-2016	107.7	59.1	166.9	27.0	23.2	13.1	4.1
Mar-2016	87.3	57.6	144.9	30.2	26.1	18.0	3.5
Apr-2016	85.5	58.8	144.4	33.9	21.1	12.5	5.9
May-2016	64.1	52.9	116.9	37.7	24.8	17.2	5.1
Jun-2016	69.8	46.7	116.6	25.7	20.1	12.5	6.9
Jul-2016	68.4	43.3	111.7	23.7	17.8	10.3	4.7
Aug-2016	63.4	44.7	108.1	26.8	21.2	11.9	3.6
Sep-2016	87.6	52.3	139.9	21.0	22.5	11.8	2.8

B.2.2 Registration of the 'Air Quality' example data set

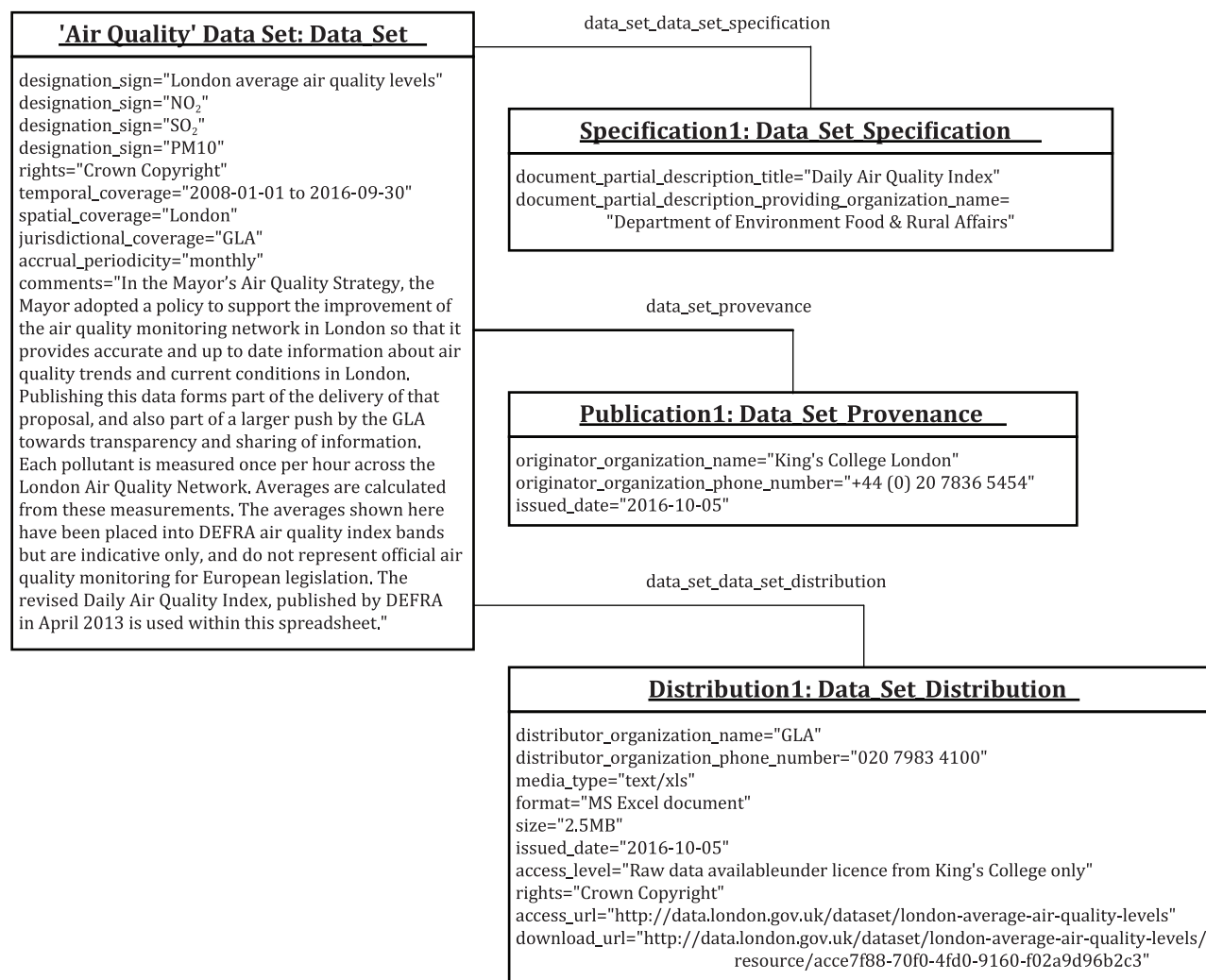Figure B.1 shows the registration of the example data set using a UML object diagram.

Figure B.1 — Registration of the 'Air Quality' example data set

B.3 The 'World War II Weather Data' data set example

B.3.1 Overview of the 'World War II Weather Data' example

This example concerns a subset of some archive data published by the US National Centers for Environmental Information which shows daily World War II era weather data from 162 land-based weather stations located outside of the United States of America. These weather stations are in countries from Algeria to the United Kingdom. The data set was accessed on 22 February 2017 from:

https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/world-war-ii-era-data.

The subset selected for this example concentrates on the data for weather station number 16407 at Simiutak, Greenland (Latitude, 6041N, Longitude 04632W, Elevation 62 metres) for the month of December 1943.

Table B.3 — World War II Weather Data' example – extract for station 16407 for December 1943

STA	YR	MO	DA	PRCP	MAX	MIN	MEA	SNF
16407	43	12	01	0.00	20	17	18	1.0
16407	43	12	02	1.78	40	20	30	1.1

Table B.3 *(continued)*

STA	YR	MO	DA	PRCP	MAX	MIN	MEA	SNF
16407	43	12	03	2.12	40	28	34	0.5
16407	43	12	04	0.00	28	20	24	0.2
16407	43	12	05	0.00	20	18	19	0.2
16407	43	12	06	0.00	26	18	22	0.2
16407	43	12	07	0.37	40	28	34	0.3
16407	43	12	08	0.04	32	27	30	0.7
16407	43	12	09	0.03	27	18	22	0.7
16407	43	12	10	0.00	18	15	16	0.8
16407	43	12	11	0.00	44	15	30	0.7
16407	43	12	12	0.00	43	36	40	0.3
16407	43	12	13	0.03	37	32	34	0.2
16407	43	12	14	T	32	28	30	0.3
16407	43	12	15	0.00	29	22	26	0.2
16407	43	12	16	0.00	24	19	22	0.2
16407	43	12	17	T	26	20	23	0.2
16407	43	12	18	0.00	20	12	16	0.2
16407	43	12	19	0.00	15	10	12	0.2
16407	43	12	20	0.00	14	10	12	0.2
16407	43	12	21	T	33	13	23	0.2
16407	43	12	22	0.40	36	23	30	0.6
16407	43	12	23	0.09	25	22	24	0.9
16407	43	12	24	0.50	24	21	22	1.1
16407	43	12	25	0.00	21	17	19	1.3
16407	43	12	26	0.17	20	17	18	1.3
16407	43	12	27	0.29	25	17	21	1.8
16407	43	12	28	0.05	23	14	18	0.8
16407	43	12	29	0.10	40	14	27	0.8
16407	43	12	30	0.03	40	26	33	1.0
16407	43	12	31	0.06	28	22	25	1.1

where:

— "STA" = Station number

— "YR" = Year

— "MO" = Month

— "DA" = Day

— "PRCP" = 24-hour precipitation measured in inches and hundredths

— "MAX" = Maximum temperature measured in degrees Fahrenheit

— "MIN" = Minimum temperature measured in degrees Fahrenheit

— "MEA" = Mean temperature measured in degrees Fahrenheit

— "SNF" = Snowfall (includes ice pellets) measured in inches and tenths

27

B.3.2 Registration of the 'World War II Weather Data' example data set

This example shows the use of the *data_set_hierarchy* association (5.1.4.10) to register the following data sets:

— the full data set;

— the subset data set for weather station16407 at Simiutak, Greenland for December 1943 (as shown in Table B.3);

— a further subset showing the data for weather station16407 for 5 December 1943 (which represents a single row from Table B.3);

— a further subset showing a single value from Table B.3 – the snowfall in inches and tenths recorded for weather station16407 for 5 December 1943 ("0.2").

The UML object diagram in Figure B.2 shows the registration of the full 'World war II Weather Data' data set. Figure B.3 shows the subsequent decomposition of this data set.
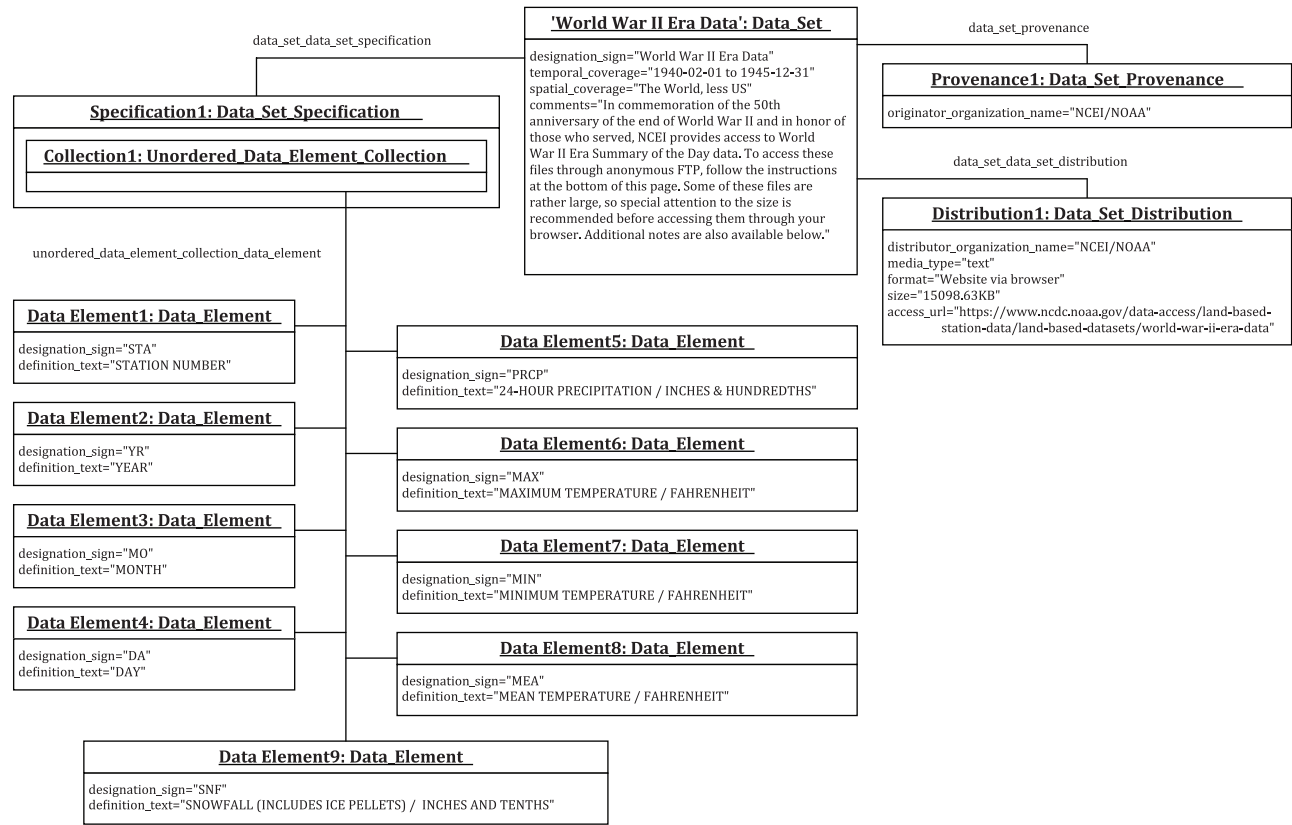

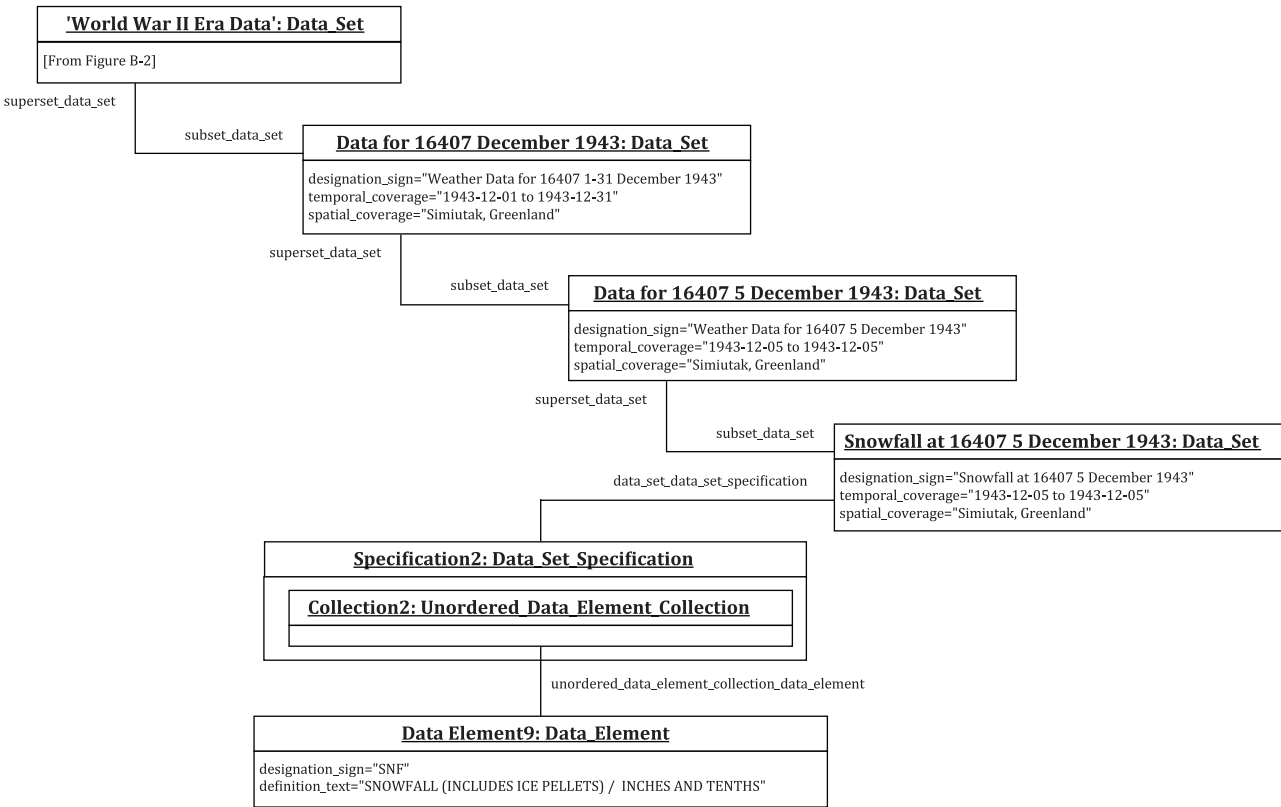**Figure B.2 — Registration of the full 'World War II Era Weather Data' data set**

**Figure B.3 — Registration of the decomposition of the 'World War II Era Weather Data'
to show the data for snowfall at weather station16407 on 5 December 1943**

Annex C
(informative)

A complete view of the concept of "provenance"

The concept of provenance as described in the Introduction and at 3.9 can be considered to be incomplete. A complete view of provenance should include coverage of the six main interrogatories of "what", "who", "when", "where", "why" and "how".

This document provides coverage of the six main interrogatories, as follows:

(a) coverage of "what":

— this is the data in the data set itself.

(b) coverage of "who":

— the value of the **originator** attribute in the **Data_Set_Provenance** (7.2.2.6) class provides the identification of who originated the data set;

— the value of the **owner** attribute in the **Data_Set_Provenance** (7.2.2.6) class provides the identification of who claims ownership of the data set.

(c) coverage of "when":

— the value of the **issued_date** attribute in the **Data_Set_Provenance** (7.2.2.6) class indicates when the data set was issued;

— the values of the **temporal_coverage** attributes in the **Data_Set** (7.2.2.3) class indicate the temporal coverage of the data set;

— the value of the **accrual_periodicity** attribute in the **Data_Set** (7.2.2.3) class indicates the accrual periodicity of the data set.

(d) coverage of "where":

— the value of the **originator** attribute in the **Data_Set_Provenance** (7.2.2.6) class is an instance of the type **Contact** (from ISO/IEC 11179-3), representing a contact that shall be linked to an organization, an instance of the **Organization** class (also from ISO/IEC 11179-3). The **Organization** class has an attribute of **mail_address**, and value of this attribute indicates where the data set was issued;

— the value of the **spatial_coverage** attribute in the **Data_Set** (7.2.2.3) class indicates the spatial coverage of the data set;

— the value of the **jurisdictional_coverage** attribute in the **Data_Set** (7.2.2.3) class indicates the jurisdictional coverage of the data set.

(e) coverage of "why":

— each instance of the **Data_Set** (7.2.2.3) class can be linked to zero, one or more instances of the **Context** (7.2.2.2) class, each of which can provide an indication of the purpose of the data set;

— each instance of the **Data_Set** (7.2.2.3) class can be linked to zero, one or more instances of the **Data_Set_Assessment** (7.2.2.7) class, each of which can indicate why a data set should be used in a particular context;

— the value of the **comment** attribute in the **Data_Set** (7.2.2.3) class can also be used to explain why a data set has been created (see the example in B.2).

(f) coverage of "how":

— the value of the **generation_type** attribute in the **Data_Set_Provenance** (7.2.2.6) class indicates how the data set was generated;

— the value of the **media_type** attribute in the **Data_Set_Distribution** (7.2.2.4.5) class indicates how the data set is made available.

(g) additional coverage:

— each instance of the **Data_Set** (7.2.2.3) class can be linked to zero, one or more instances of the **Data_Set_Specification** (7.2.2.8) class, each of which could be an instance of the **Reference_Document** class (specified in ISO/IEC 11179-3), an instance of the **Model** class (specified in ISO/IEC 11179-35), or a collection of instances of the **Data_Element** class (specified in ISO/IEC 11179-31). Any of these could provide further evidence of the provenance ("what", "who", "when", "where", "why", "how" or some combination) of the data set.

Bibliography

[1]   W3C Data Catalog Vocabulary (DCAT) (http://www.w3.org/TR/vocab-dcat/) (accessed 5 August 2022)

[2]   W3C The PROV Ontology (Prov-O) (http://www.w3.org/TR/prov-o/) (accessed 5 August 2022)

[3]   W3C The PROV Data Model (Prov-DM) (http://www.w3.org/TR/prov-dm/) (accessed 5 August 2022)

[4]   Project Open Data Metadata Schema v1.1 (https://project-open-data.cio.gov/v1.1/schema) (accessed 5 August 2022)

[5]   Apache Avro™ 1.11.0 Specification (https://avro.apache.org/docs/current/spec.html) (accessed 5 August 2022)

[6]   Java Script Object Notation (JSON) Data Interchange format (https://datatracker.ietf.org/doc/html/rfc8259) (accessed 5 August 2022)

[7]   RDF 1.1 Concepts and Abstract Syntax (https://www.w3.org/TR/rdf11-concepts/) (accessed 5 August 2022)

[8]   Extensible Markup Language (XML) 1.0 (Fifth Edition) (https://www.w3.org/TR/REC-xml/) (accessed 5 August 2022)

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

[10]   ISO/IEC 19505-1:2012, *Information technology — Object Management Group Unified Modeling Language (OMG UML) — Part 1: Infrastructure*

[11]   ISO/IEC 19505-2:2012, *Information technology — Object Management Group Unified Modeling Language (OMG UML) — Part 2: Superstructure*

[12]   MEDIA TYPES I.A.N.A., https://www.iana.org/assignments/media-types/) (accessed 5 August 2022)

[13]   ISO 8000-8, *Data quality — Part 8: Information and data quality: Concepts and measuring*

[14]   ISO 19157, *Geographic information — Data quality*

[15]   ISO 25012, *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model*

[16]   ISO/IEC 19763-12:2015, *Information technology — Metamodel framework for interoperability (MFI) — Part 12: Metamodel for information model registration*

[17]   ISO/IEC 2382:2015, *Information technology — Vocabulary*

[18]   ISO 2146:2010, *Information and documentation — Registry services for libraries and related organizations*

[19]   ISO/IEC 11179-1, *Information technology — Metadata registries (MDR) — Part 1: Framework*

ICS 35.040.50

Price based on 32 pages