FINAL DRAFT
International Standard

ISO/IEC
FDIS
11179-34

Information technology — Metadata registries (MDR) —

Part 34: Metamodel for computable data registration

ISO/IEC JTC **1**/SC **32**

Secretariat: **ANSI**

Voting begins on:
2024-02-22

Voting terminates on:
2024-04-18

Reference number
ISO/IEC FDIS 11179-34:2024(en)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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 document is part of the fourth modularization of ISO/IEC 11179. This document brings into ISO/IEC 11179 the ability to register information about computable data.

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

Significant scientific discoveries are increasingly achieved through complex and distributed computations and data analyses. These computations and analyses always involve processing files through a series of steps and transformations, usually called a pipeline or a workflow.

Data pipelines typically use multiple pieces of software, each of which typically has multiple versions available, multiple input parameters, multiple outputs, and possibly platform-specific configurations. As with experimental parameters in a laboratory protocol, small changes in computational parameters can have a large impact on the scientific validity of the results.

To reproduce and verify scientific discoveries, details of pipelines need to be documented and shared, including the protocol, procedures, or metadata associated with them. The more analysis steps and the more complicated a pipeline, the greater the need for a standardized mechanism of communication. A detailed communication helps ensure responsibility, reproducibility, and the ability to verify protocol, track provenance information, and promote interoperability.

This document is inspired by the IEEE 2791-2020[1], which aims to improve communication of bioinformatics protocols and data to facilitate bioinformatics workflow related exchange and communication between regulatory agencies, pharmaceutical companies, bioinformatics platform providers and researchers. Although IEEE 2791-2020 has a bioinformatics background and application areas, the concepts and methods it expressed are applicable to a broader field of scientific research. A mapping table is included in Annex A, showing the relationship between domains and fields in IEEE 2791-2020 and classes and attributes in this document.

ISO/IEC 11179-3 specifies the structure of a Metadata Registry (MDR) and provides a metamodel for registry common facilities. That 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 computable data. Registration of metadata about computable data are like a manifest describing all details related to input files, output files, and the pipeline used to process these files. The intent is to facilitate efficient communication and interoperability among different platforms, industries, scientists, and regulators and to improve reproducibility and replicability.

In Clauses 6 and 7, this document uses **bold** font to highlight terms which represent metadata objects specified by the metamodel.

EXAMPLE **Computable_Data** (see 7.2.2.1) is a class each instance of which models computable data.

Information technology — Metadata registries (MDR) —

Part 34:
Metamodel for computable data 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 that describe computable data can be registered.

The specification in this document, together with the relevant clauses of the specification in ISO/IEC 11179-3, provides the ability to record metadata about computable data.

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, *Information technology — Metadata registries (MDR) — Part 3: Metamodel for registry common facilities*

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 terminology 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
computable data
data that are computed for the purpose of making the data reproducible

3.2
pipeline
full set of logically connected steps needed to transform input data into a result, for which one or more outputs of one data transformation step can be the input of one or more subsequent steps

3.3
computation step
individual tool (or a well defined and reusable script) used in a *pipeline* (3.2) that can be executed sequentially or in parallel with other computation steps

3.4
computation execution environment
information of the environment needed to execute a *computation step* (3.3) to the extent to which consistent results are obtained when produced repeatedly, including platform, deployment, software configuration and running applications, etc.

4 Abbreviated terms

JSON JavaScript Object Notation

ORCID Open Researcher and Contributor ID(see Reference [2])

PAV Provenance, Authoring and Versioning (see Reference [3])

UML Unified Modeling Language (see References [4] and [5])

URI Uniform Resource Identifier

URL Uniform Resource Locator

URN Uniform Resource Name

5 Conformance

5.1 Overview of conformance

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

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;

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 can 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 of this document, or to specific features within that clause. Clause 7 is also dependent upon one or more clauses of ISO/IEC 11179-3, so conformance to all or part of Clause 7 shall be understood to imply conformance also to relevant provisions specified in one or more of the clauses in ISO/IEC 11179-3.

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.

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

— **Computable data Registry with mapping**: Implements clause 7 of this document, 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 11179-34:2024 Computable data Registry;

— ISO/IEC 11179-34:2024 Computable data 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 11179-34:2024 Computable data Registry.

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, 9.4.6.3 and ISO/IEC 11179-6:2023, 4.4).

6 Relationship to ISO/IEC 11179-3

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 both.

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. Many of the classes specified in this document inherit from *Item*, which is specified in ISO/IEC 11179-3:2023, 6.4.2.1. As *Items*, instances of these classes 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 to describe the metamodel. Both 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 B.

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.

6.4 Package dependencies

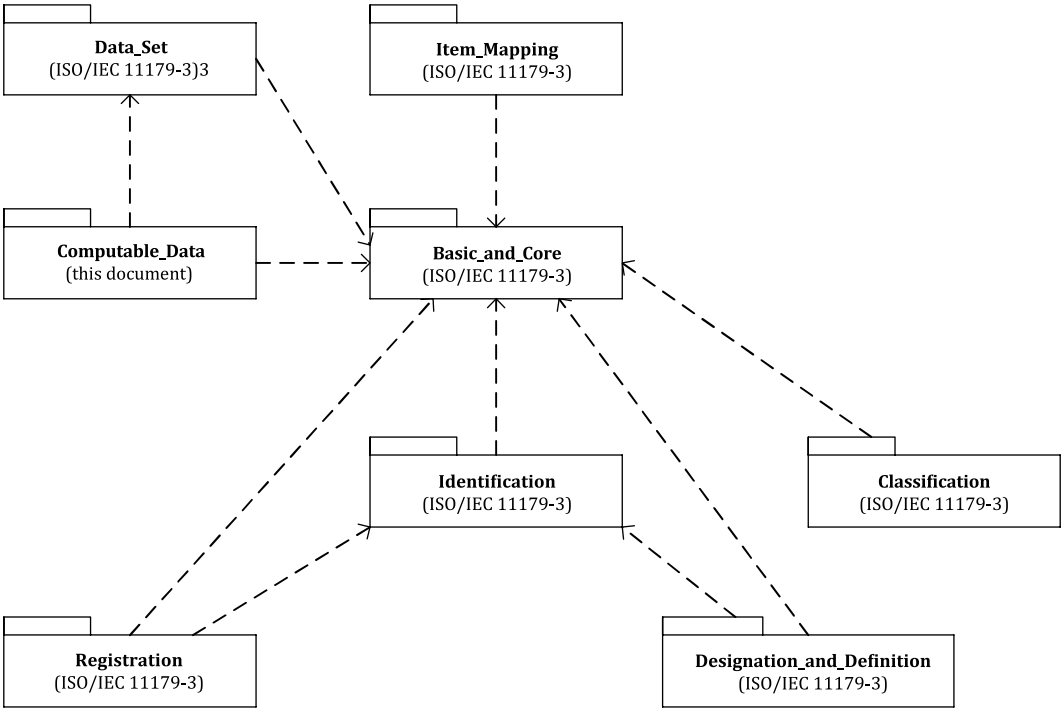

Figure 1 — Package dependencies

Figure 1 illustrates the dependencies among the packages. The Computable_Data package is specified in this document. All the other packages are specified in ISO/IEC 11179-3 and ISO/IEC 11179-33.

The lines in Figure 1 illustrate dependencies in the direction of the arrow. 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. **Computable_Data** (see 7.2.2.1) in the Computable_Data package is subclassed from the **Item** class in the Basic and Core package (ISO/IEC 11179-3:2023, 6.4.2.1);

b) relationship between classes, e.g. **Registered_Item** in the Registration package (ISO/IEC 11179-3:2023, 9.4.1) has a relationship with **Reference_Document** in the Basic and Core package (ISO/IEC 11179-3:2023, 6.3.8);

c) some attributes use a predefined datatype or a class from another package as a datatype, e.g. the **supporting_document** attribute of the **Supporting_Document** (see 7.2.2.3) class in the Computable_Data package (see Clause 7) uses the **Reference_Document** class of the Basic and Core package (see ISO/IEC 11179-3:2023, 6.3.8) as a datatype.

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

7 Computable_Data package

7.1 Overview of the Computable_Data package

The **Computable_Data** package consists of a single metamodel region, the Computable_Data metamodel region.

7.2 Computable_Data metamodel region

7.2.1 Overview of the Computable_Data metamodel region

Figure 2 shows the framework of the metamodel for the registration of metadata about computable data, Figure 3 and Figure 4 show the detailed information of this metamodel.

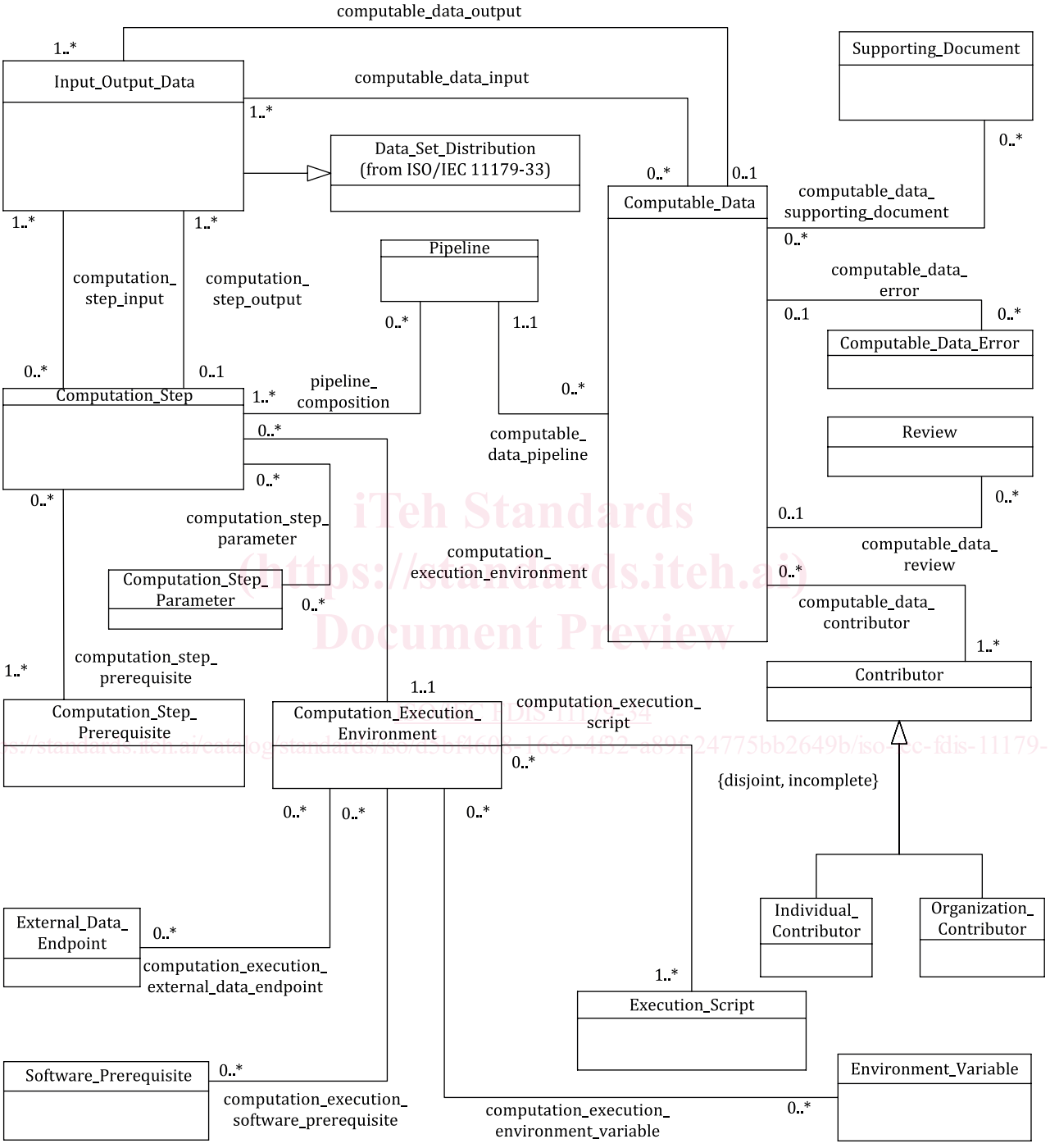

Figure 2 — Framework of computable data metamodel region