**BSI Standards Publication**

# Information technology — Metadata registries (MDR)

Part 7: Metamodel for data set registration

**bsi.**

## National foreword

This British Standard is the UK implementation of ISO/IEC 11179-7:2019.

The UK participation in its preparation was entrusted to Technical Committee IST/40, Data management and interchange.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2019
Published by BSI Standards Limited 2019

ISBN 978 0 580 97989 7

ICS 35.040.50

 **Compliance with a British Standard cannot confer immunity from legal obligations.**

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 December 2019.

### Amendments/corrigenda issued since publication

| Date | Text affected |
|------|---------------|

INTERNATIONAL
ISO/IEC

First edition
2019-12

# Information technology — Metadata registries (MDR) —

Part 7:
**Metamodel for data set registration**

# 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.

INTERNATIONAL STANDARD                                                        ISO/IEC 11179-7:2019(E)

# 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.