TECHNICAL REPORT

ISO/IEC TR 19583-1

First edition
2019-07

Information technology — Concepts and usage of metadata —

Part 1: Metadata concepts

Technologies de l'information — Concepts et utilisation des métadonnées —

Partie 1: Concepts liés aux métadonnées

Reference number
ISO/IEC TR 19583-1:2019(E)

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

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

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

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

A list of all parts in the ISO/IEC 19583 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

This document describes the concept of metadata, particularly in respect of its use within the data management speciality in information technology.

The ISO/IEC 11179[2] and ISO/IEC 19763[4] series describe the structure for registering information about metadata that is used and/or held elsewhere.

The ISO/IEC 11179 series defines metadata as “data that defines and describes other data”. There are, however, many other definitions of metadata that are used more generally, for example, the US National Information Standards Organization (NISO) defines metadata as "structured information which describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”.

Metadata is, therefore, just data, but data which has the specific purpose of defining or describing other data. Metadata is normally used within a particular context, which is the set of circumstances, purposes or perspectives within which any particular item of data is used as metadata. Metadata can, therefore, be considered to be data about data within some context.

The definitions above, by themselves, do not say how metadata arises, where it comes from, how it is used, or how it is managed (although the ISO/IEC 11179 series describes the facilities for registering and managing structured metadata). For those reasons, this document has been developed to provide a broader view of metadata and the associated concept of the metamodel.

These concepts of metadata and metamodels are important when trying to understand exactly what is being registered within the registries whose structure is specified in the ISO/IEC 11179 and ISO/IEC 19763 series.

Information technology — Concepts and usage of metadata —

Part 1: Metadata concepts

1 Scope

This document describes the basic concept of metadata, and its relationship to both data and metamodels.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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/

4 Introduction to metadata

Metadata is defined as “data that defines and describes other data”. This is a very broad definition of metadata leaving room for confusion. It is often said that one person’s metadata is another person’s data. Whether any piece of data is seen as metadata or just data depends on the context. These contexts can be classified into three distinct groups:

— structural metadata: the metadata used by those responsible for the management of data in information systems to describe the ‘containers’ of data, for example, the tables and columns in a database managed using the SQL database language;

— descriptive metadata: the metadata used for the discovery and identification of content, such as by librarians, and the metadata that helps to further describe other data, such as metadata that a scientist has observed about continuous or systematically produced data;

— administrative metadata: the metadata associated with data values, such as metadata describing when and who created the data, who can edit and manage the data, and any other information about the data that is deemed useful, including metadata that describes multimedia data.

The focus of this document is the first of these groups: the structural metadata used within data management. The use of descriptive metadata is explained in Annex A and the use of administrative metadata is explained in Annex B.

For any data to be useful or shareable the meaning of the data (the semantics), the data type and format of the data (the syntax) and the relationship of the data to other data (the structure) must be known. All of this information about data is metadata.

Metadata is independent of the systems that produce the data. Metadata is usually defined before systems are built, either as part of the systems development of an individual system or as part of an

enterprise-wide data management initiative. However, metadata can be recorded after the data has been created in less formal systems as part of a data documentation initiative. The only significant difference between single-system metadata and enterprise-wide metadata is the scope of the metadata, although there might be differences around the degree of formality applied to its creation.

5 The use of structural metadata in data management

The traditional data management view of metadata is that it describes the types of data stored in a database and also describes how that data is to be managed. Examples of such metadata are:

— table and column definitions for a database schema managed by an SQL database management system;

— definitions of any constraints used for validating the data to be placed in the database;

— rules for accessing the data in the database;

— rules for maintaining the quality of the data in the database;

— predicted volumes for the data in the database.

A more comprehensive view of metadata within the data management community is that metadata is defined very early in the systems development lifecycle[6]. Conceptual data models and the data definitions derived from them can, therefore, be considered as metadata. The relationships shown on a conceptual data model can also be considered as metadata.

At the enterprise level, other information about data can be viewed as metadata. Details of the ownership and source of data definitions can be viewed as metadata, as can any other information that helps business users and system developers with understanding what data is recorded in the enterprise's databases, and where it is recorded.

Metadata may be held on paper or electronically, or both. In addition to the diagrams of the data models for the system, paper-based metadata can include glossaries of business terms that support the use of the data, descriptions of the information systems owned by the enterprise, and descriptions of the data held by each system. Metadata that is held electronically may be stored in the information technology systems, for example, in the system tables, where this metadata supports the creation and management of the data held in the individual systems. Metadata may also be held electronically within a data warehouse to describe the schemas of the operational systems that feed the data warehouse to assist with the transformation and loading of that data into the data warehouse.

Metadata may also be held electronically in stand-alone systems to support interoperability by enabling the common understanding of data that is shared between separate systems. These stand-alone systems are often called data dictionaries, repositories or registries. Such stand-alone systems will each have its own database. The structure of that database will be represented in a series of data models that together describe what information about the metadata is to be held and how it is to be held. The relationships between the information may also be described in the models.

These data models can be produced at three levels:

— The conceptual model, or computation independent model, is a model that specifies ***what*** information is to be held by the system (i.e., the data dictionary, repository or registry). Note that with a conceptual model there is no assumption that the final system will be using any information technology; its implementation could be paper-based.

— The logical model, or platform independent model, is the model that begins to consider ***how*** the information is to be held in an information technology-based system. For example, a logical model can be developed on the assumption that the system will have a database managed using the SQL database language.

— The physical model, or platform specific model, is the model that specifies ***how*** the information is to be held in the particular implementation selected for the system. For example, at this level consideration should be given to such details as the naming and allocation of tables to tablespaces.

Each of these models (conceptual, logical and physical) for a data dictionary, repository or registry is known as a metamodel – a model of a model.

Metamodels need not only be used to describe the requirements to hold information about metadata. It is often useful for repositories and registries to hold information about other resources that are important to an enterprise, such as the processes supported by an enterprise's systems, information about the roles and goals of the user of those systems, and information about the forms used to collect the data to be entered into those systems. This required information (including any rules and constraints for describing these resources) will also be represented in one or more metamodels.

6 The relationship between data, metadata and metamodels

A convenient explanation of the relationship between data, metadata and metamodels is provided in ISO/IEC 10027:1990, Clause 6[1], where four data levels are described. The purpose of these four data levels is to make it possible to extend the types of data that can be held in the Information Resource Dictionary (IRD), the name given in ISO/IEC 10027 to a data dictionary, a repository or a registry.

The four data levels described are:

— IRD Definition Schema Level;

— IRD Definition Level;

— IRD Level;

— Application Level.

These levels are illustrated in Figure 1.

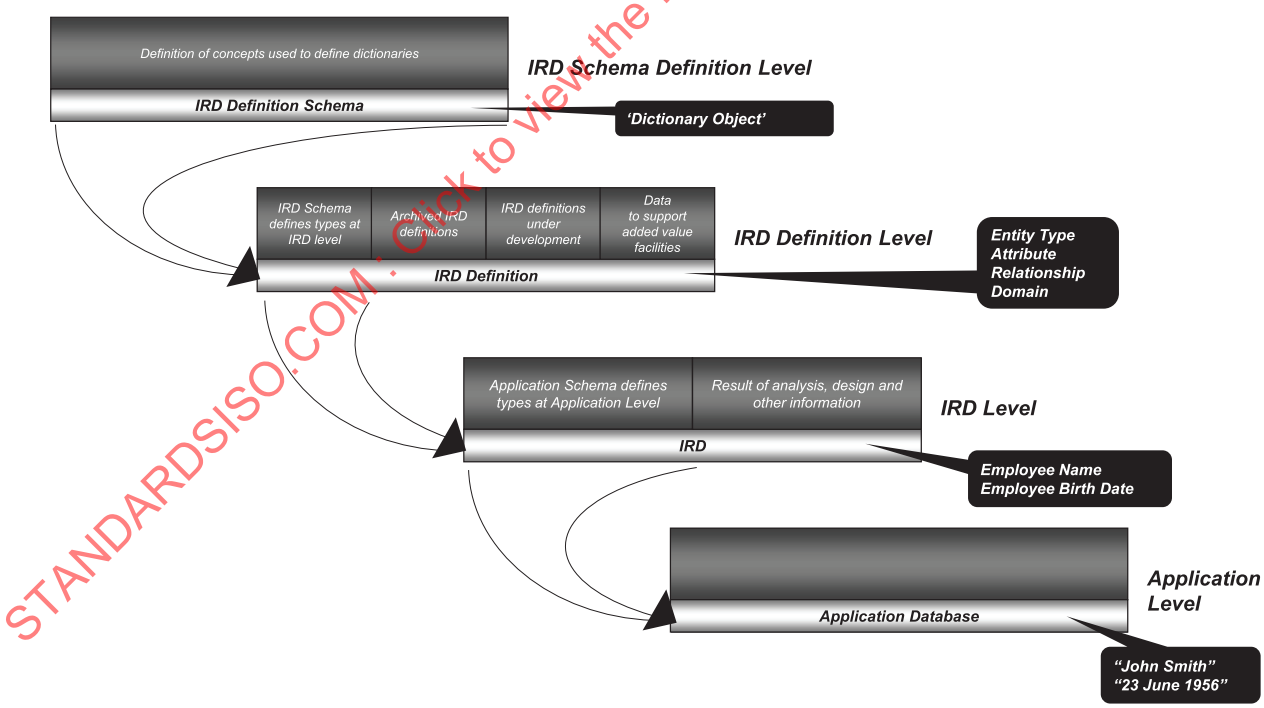

Figure 1 — Four data levels

The Application Level is the level on which instances of business data are recorded in the Application Database. For example, the data about specific instances of employee will be recorded at the Application Level, such as the data about an employee with the name "John Smith" and with a date of birth of "23 June 1956". The Application Level exists, therefore, in the operational information system and not in the Information Resource Dictionary.

The IRD Level describes two types of information about the Application Level.

The first type of information at the IRD Level is the specification of results of the analysis and design, and other information of the systems that are to hold the business data. For example, it is at this level that we would identify that, for each employee, we need to hold their name (Employee Name) and their date of birth (Employee Birth Date). Alongside this basic information we may also have formal definitions for the concept of Employee Name and the concept of Employee Birth Date (the semantics) and the data type and format associated with each concept (the syntax). We would probably identify the fact that Employee Name and Employee Birth Date are attributes of the entity type (or object class) EMPLOYEE (the structure).

The second type of information at the IRD Level is the actual schema definitions for the systems that will hold the business data, such as the definition of an EMPLOYEE table whose columns include Employee_Name and Employee_Birth_Date.

The contents of the IRD Level – instance of the schema definition – are metadata because they describe instances of application data.

The IRD Definition Level is a set of containers for information at the IRD Level. The definition of these containers is also metadata because it describes the entities and attributes that need to be in the schema at the IRD Level. The design of the IRD Definition Level is implemented using a series of metamodels. A conceptual metamodel will define what information is to be recorded at the IRD Definition Level whilst the logical and physical models will define how that information is to be recorded. For example, it is at the IRD Definition Level that we recognize that we have modelling concepts such as: Employee (Entity Type, Object Class), Name (Attribute), Relationship (or Association); and implementation concepts such as: Table, Domain, Schema. At the IRD Definition Level we also need to be able to record any archived definitions of the schema at the IRD Level, any definitions at the IRD Level that are under development or any other information that will describe support facilities that will add value to the business. The metamodels at the IRD Definition Level must support all of these requirements.

Metadata registries are systems that exist at the IRD Definition Level only. The metamodel for a Metadata Registry, as specified in ISO/IEC 11179, also needs to specify information requirements to enable the registration of metadata and the contexts for which the metadata concerned is appropriate. The metamodel for a Metadata Registry that is extended by the facilities specified in ISO/IEC 19763 also needs to provide for the recording of information about ontologies, information models, process models, role and goal models, models of web services and/or form definitions. There has been a distinction made between ISO/IEC 11179 metadata models that describe data, and ISO/IEC 19763 metamodels that describe models. This distinction may disappear over time.

The IRD Definition Schema Level specifies the concepts to be used at the IRD Definition Level. In the preparation of ISO/IEC 11179 and ISO/IEC 19763 it was decided to use the Unified Modeling Language™ (UML) class diagram notation, but an entity-relationship notation could also have been used.

Annex A
(informative)

Descriptive metadata

A.1 General

Increasingly, the term metadata is also being used to describe data that describes the content of documents held in libraries and other archives and the content of webpages. Descriptive metadata is data that exists within a database to provide additional information to:

— find and interpret other data;

— record observations about other data held in the database.

A.2 Document metadata

A typical example of this use of the term metadata to describe data that describes the content of documents held in libraries and other archives is the Dublin Core Metadata Element Set[3], often informally known as the Dublin Core ('Dublin' is Dublin, Ohio, USA, the home of the Online Computer Library Center).

The Dublin Core is a standard for describing information resources – video, sound, image, text and webpages – that may be used in many different contexts. The first version of the Dublin Core was designed for simple resource discovery. It has 15 metadata elements, each of which is optional and may be repeated:

— title: a name by which the resource is known;

— creator: a person, an organisation or a service;

— subject;

— description: which may include an abstract table of contents, reference to a graphical representation of the content or a free-text account of the content;

— publisher;

— contributor;

— date;

— type: whether the resource is a moving image, a still image, sound, text, etc.;

— format;

— identifier: for example, a uniform resource identifier (URI) or an international standard book number (ISBN);

— source;

— language;

— relation: any related resources;

— coverage: spatial, temporal or jurisdiction;

— rights: for example, intellectual property rights (IPR), copyright and various property rights.

Another version of the Dublin Core, the Qualified Dublin Core, has been developed. There are three additional elements (audience, provenance and rightsHolder). Some of the simple Dublin Core elements may also be qualified; for example, the date element may be qualified to be available, created, dateAccepted, dateCopyrighted, dateSubmitted, issued, modified or valid.

The Dublin Core is not the only standard for content management. Amongst the many others available are the Standard for Learning Object Metadata[5] published by the Institute of Electrical and Electronics Engineers (IEEE), and the Agricultural Metadata Element Set (AgMES)[7] from the Food and Agriculture Organisation (FAO) of the United Nations.

The description of the content of web pages is achieved through the embedding of metadata in the HTML (HyperText Markup Language) using metatags. These metatags are used to provide details such as a general description of the content of the webpage, appropriate keywords, author of the webpage, and the date the webpage was last updated.

These metatags are then used by browsers to help with the display of the webpages and by search engines to find appropriate webpages.

A.3 Observations about data

Some descriptive metadata provides observations about other data. An example of this use of descriptive metadata, is where a pathologist has made and wishes to record observations about the image on a pathology slide. These observations might include, for example that "the neoplasm was poorly differentiated", or that "there was considerable necrosis in the sample". Sometimes these observations may themselves conform to a data model, that itself may have conceptual and logical implementations – the UK Royal College of Pathologists specifies a large set of conceptual proformas for the recording of metadata about a cancer pathology sample as PDF documents alongside detailed descriptions of pathology practice and the justification for any recommendations on the interpretation of samples; the American College of Pathologists communicates its proformas in a machine readable form.

Other observations about data may be used to understand if the data in a dataset is usable for a particular context: statistical data such as the frequencies of categorical values; the percentage completion of data in columns; the distribution of values in columns; the number of records of a particular kind all may be used to decide if a dataset, or portion thereof, is of use for some defined purpose. The use of this type of descriptive data for the cataloguing and discovery of data included in a data set is addressed in ISO/IEC11179-7[1)].

1) Under preparation. Stage at time of publication: ISO/IEC DIS 11179-7:2019.

