TECHNICAL REPORT

ISO/IEC TR 11179-2

First edition
2019-04

Information technology — Metadata registries (MDR) —

Part 2: Classification

Technologies de l'information — Registres de métadonnées (RM) — Partie 2: Classification

Reference number
ISO/IEC TR 11179-2:2019(E)

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

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

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

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

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

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document focuses on the part of the metadata registry (MDR) model called the *classification region* (see ISO/IEC 11179-3:2013, 9.2). The classification region permits the registration and administration of all or part of a *classification scheme*. A classification scheme can be used to classify a *Classifiable Item* (see ISO/IEC 11179-3:2013, 9.2.2.1), which is a type that can be applied to any metadata item in a metadata registry.

There are many efforts underway to devise classification schemes and to use the schemes to build and populate classification structures for organizing information resources. Classification schemes can be used to discover information pertaining to topics of interest. For the purpose of this document, the following are all considered types of classification schemes of varying discriminatory power: tags, keywords, lists of categories, hierarchies, thesauri, taxonomies and ontologies. These classification schemes have potentially great utility for organizing objects in an MDR.

When applied to classifiable items in an MDR, the classification schemes covered in this document have utility for:

— deriving and formulating other administered items;

— ensuring appropriate attribute and attribute-value inheritance;

— deriving names from a controlled vocabulary;

— disambiguating;

— recognizing superordinate, coordinate and subordinate concepts;

— recognizing relationships among items;

— deriving or clarifying the meaning of items associated with the classified items;

— assisting in the development of modularly designed names and definitions.

Each type of classification scheme mentioned above has particular strengths and weaknesses, and provides the foundation upon which particular capabilities can be built. Tags and keywords, for example, are a quick way to provide users some assistance in locating potentially useful classifiable items. A thesaurus provides a more structured approach, arranging descriptive terms in a structure of broader, narrower and related classification categories. A taxonomy provides a classification structure that adds the power of inheritance of meaning from generalized taxa to specialized taxa. Ontologies, with associated epistemologies, can provide rich, rigorously defined structures (e.g. directed acyclic graphs with multiple inheritance) that can convey information needed by software, such as intelligent agents and mediators that are useful in the provision of intelligent information services. When classification systems are used to classify data elements, or value domains, it makes it easier for end users to interpret the data that is associated with the data elements or value domains.

An example of a classification scheme that uses external standards or controlled vocabulary in registering to a metadata registry is provided in Annex A.

Information technology — Metadata registries (MDR) —

Part 2: Classification

1 Scope

This document compliments ISO/IEC 11179-3 by describing registration of classification schemes and using them to classify registered items in an MDR. Any metadata item can be made a *Classifiable_Item* so it can be classified, which can include object classes, properties, representations, conceptual domains, value domains, data element concepts and data elements themselves.

This document does not establish a particular classification scheme as pre-eminent. Sanction of a particular taxonomic approach and/or a particular epistemology is also beyond the scope of this document. These are addressed by other standards committees and/or tend to be tailored to a particular domain of discourse. The MDR can establish its own classification schemes, and other standards committees are developing or have developed normative languages for use in classification and/or particular techniques and structures that can be accommodated by this document.

Each registration authority, as described and specified in ISO/IEC 11179-6, can classify classifiable items according to the classification schemes, structures and content that it deems appropriate. In documenting the classification aspects of classifiable items, the registration authority can use the principles, methods, procedures and attributes covered in this document.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions 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
characteristic
abstraction of a property of an *object* (3.5) or of a set of objects

Note 1 to entry: Characteristics are used for describing *concepts* (3.3).

[SOURCE: ISO 1087-1:2000, 3.2.4]

3.2
classification scheme
descriptive information for an arrangement or division of *objects* (3.5) into groups based on criteria such as *characteristics* (3.1), which the objects have in common

[SOURCE: ISO/IEC 11179-3:2013, 3.2.16]

Note 1 to entry: A classification scheme is a concept system used for classifying some objects.

3.3
concept
unit of knowledge created by a unique combination of *characteristics* (3.1)

Note 1 to entry: Concepts are not necessarily bound to particular languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.

[SOURCE: ISO 1087-1:2000, 3.2.1]

3.4
concept system
set of *concepts* (3.3) structured according to the relations among them

[SOURCE: ISO 1087-1:2000, 3.2.11]

3.5
object
anything perceivable or conceivable

Note 1 to entry: Objects may also be material (e.g. an engine, a sheet of paper, a diamond), immaterial (e.g. a conversion ratio, a project plan), or imagined (e.g. a unicorn).

[SOURCE: ISO 1087-1:2000, 3.1.1]

3.6
relation
sense in which *concepts* (3.3) can be connected, via constituent roles

EXAMPLE Causality is a relation with two constituent roles: cause and effect.

[SOURCE: ISO/IEC 11179-3:2013, 3.2.119]

3.7
hierarchical relation
relation between two concepts which can be either a generic relation or a partitive relation

[SOURCE: ISO 1087-1:2000, 3.2.20]

3.8
hierarchy
concept system in which all concepts are related in hierarchical relations that form a partial ordering

4 Meaning of classification

Classifying may refer to categorizing, the process in which ideas and objects are recognized, differentiated and understood. In general, ideas and objects can be classified by a **classification scheme**.

For the purpose of this document, a **classification scheme** defines descriptive information for an arrangement or division of objects (ideas) into groups based on criteria such as characteristics, which the objects (ideas) have in common (ISO/IEC 11179-3:2013, 3.2.16).

A classification scheme is represented as a **concept system** (ISO/IEC 11179-3:2013).

A **concept system** is a set of **concepts** structured according to the **relations** among them (ISO/IEC 11179-3:2013, 3.2.19) and inherited by the classifiable item.

A **concept** is defined as a unit of knowledge created by a unique combination of characteristics (ISO/IEC 11179-3:2013, 3.2.18).

Relation is defined as a sense in which **concepts** may be connected, via constituent roles.

Figure 1 depicts the above definitions of key words related with classification.

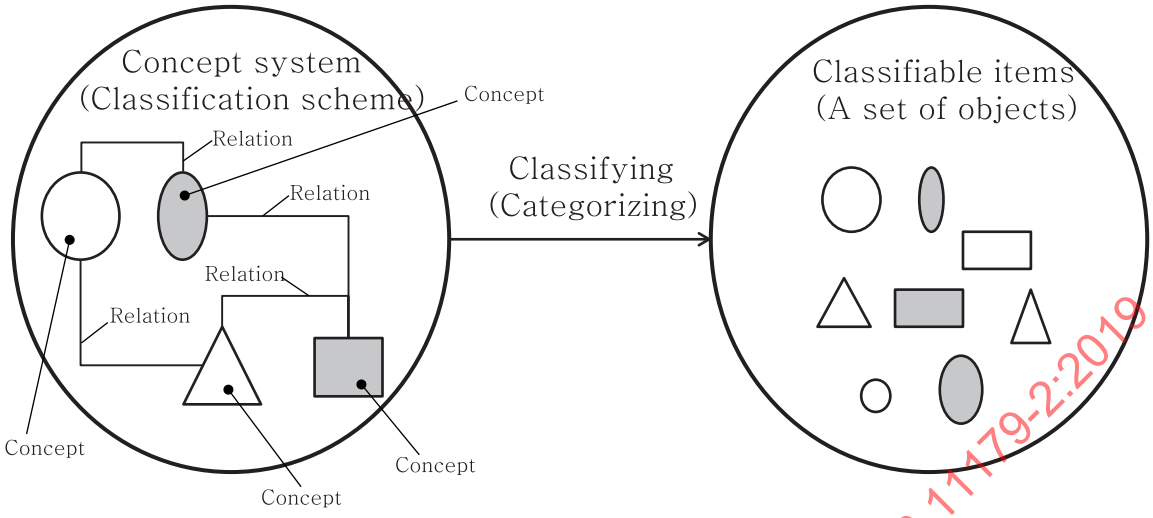

Figure 1 — Meaning of classification

5 Benefits of classification scheme

Using classification schemes for the classification of a collection of objects has many benefits. Use of one or more classification schemes is intended to provide a sound conceptual basis for the development of metadata having enhanced semantic purity and design integrity. A classification scheme: enables users to find a single object from among a large collection of objects in the metadata registry; facilitates the discovery, administration and analysis of a collection of objects; and, through inheritance, conveys semantic content that is often only incompletely specified by other attributes, such as names and definitions. It also allows a user to find objects quickly on the basis of its kind or group and makes it easier to detect duplicate objects. It conveys additional semantics (meaning) which may not be conveyed by other attributes of the object, such as its name and definition.

6 Type of classification scheme

Typical classification schemes include **controlled vocabulary**, **uncontrolled vocabulary**, **taxonomy**, **thesaurus**, and **ontology**.

Controlled vocabulary is a list of terms that have been enumerated explicitly (see Figure 2). This list is controlled by, and is available from, a controlled vocabulary registration authority. All terms in a controlled vocabulary should have an unambiguous, non-redundant definition. This is a design goal that may not be true in practice. It depends on how strict the controlled vocabulary registration authority is regarding registration of terms into a controlled vocabulary. At a minimum, the following two rules should be enforced[5]:

— If the same term is commonly used to mean different concepts in different contexts, then its name is explicitly qualified to resolve this ambiguity.

— If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term in the controlled vocabulary and the other terms are listed as synonyms or aliases.

EXAMPLE country code, language code, colour code, etc.

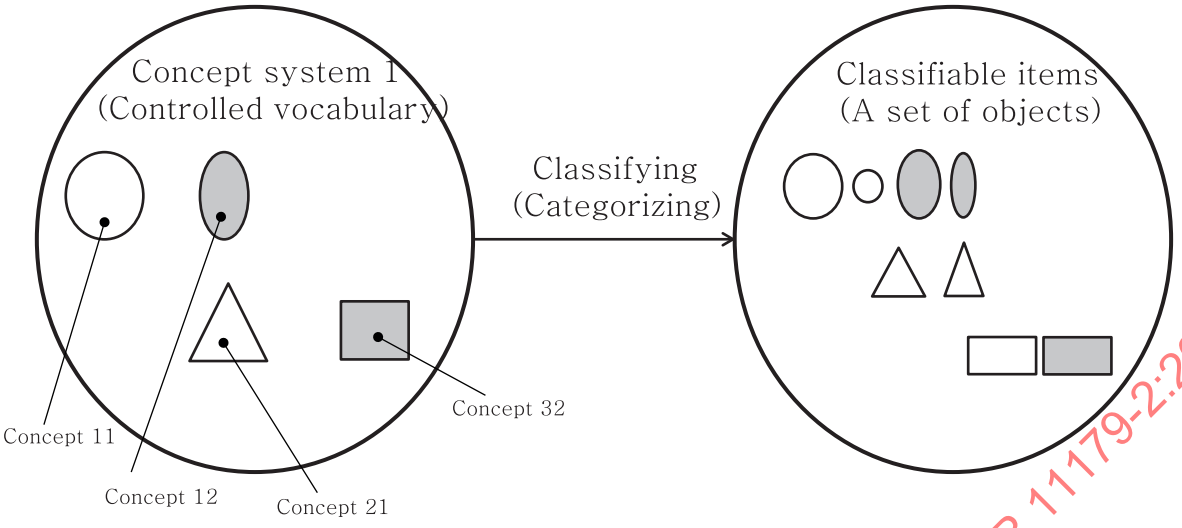

Figure 2 — Concept of classification (controlled vocabulary)

Uncontrolled vocabulary is a list of terms that have not been enumerated explicitly (see Figure 3).

EXAMPLE keywords, tag sets, etc.

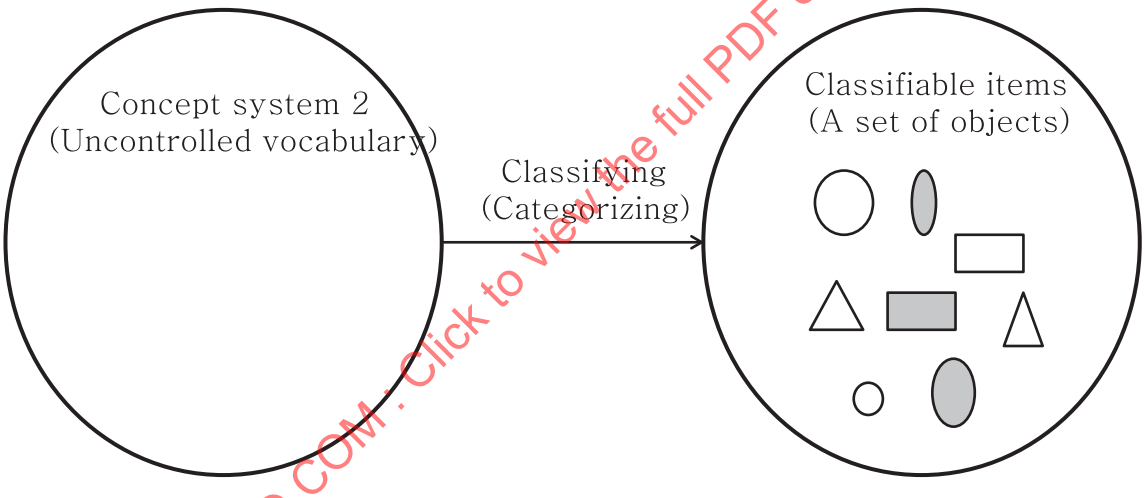

Figure 3 — Concept of classification (uncontrolled vocabulary)

Taxonomy is a collection of controlled vocabulary terms organized into a hierarchical structure (see Figure 4). Each term in taxonomy is in one or more parent-child relationships to other terms in the taxonomy. There may be different types of parent-child relationships in a taxonomy (e.g., whole-part, genus-species, type-instance), but good practice limits all parent-child relationships to a single parent to be of the same type. Some taxonomy allows poly-hierarchy, which means that a term can have multiple parents. This means that if a term appears in multiple places in taxonomy, then it is the same term. Specifically, if a term has children in one place in taxonomy, then it has the same children in every other place where it appears[5].

EXAMPLE Dewey decimal classification (DCC), zoological taxonomy, plant taxonomy, etc.

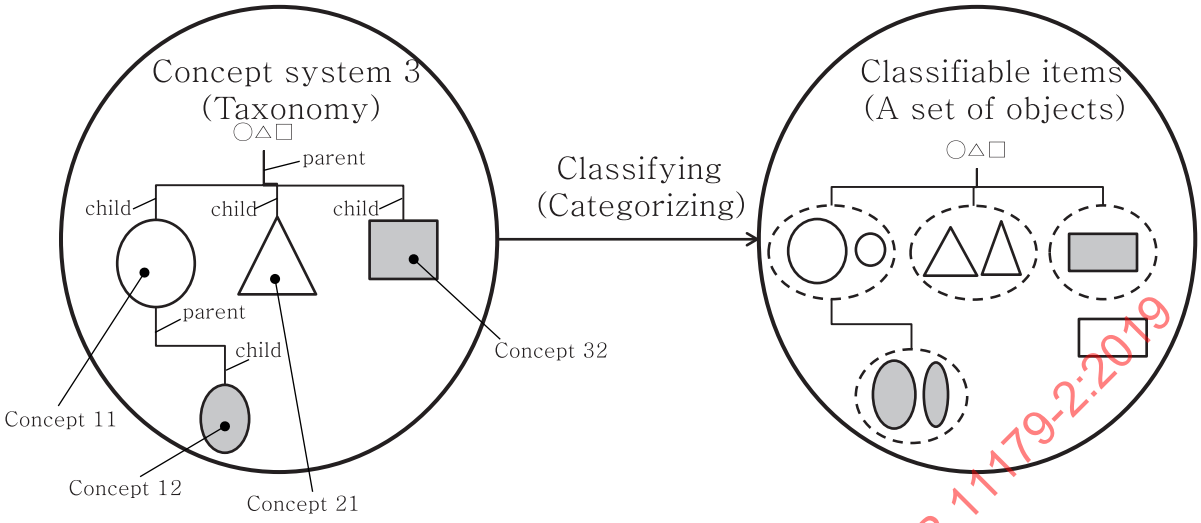

Figure 4 — Concept of classification (taxonomy)

Thesaurus is a networked collection of controlled vocabulary terms (see Figure 5). This means that a thesaurus uses associative relationships in addition to parent-child relationships[5] such as synonyms and antonyms.

EXAMPLE Medical subject heading (MeSH), NASA thesaurus, UNESCO Thesaurus, etc.

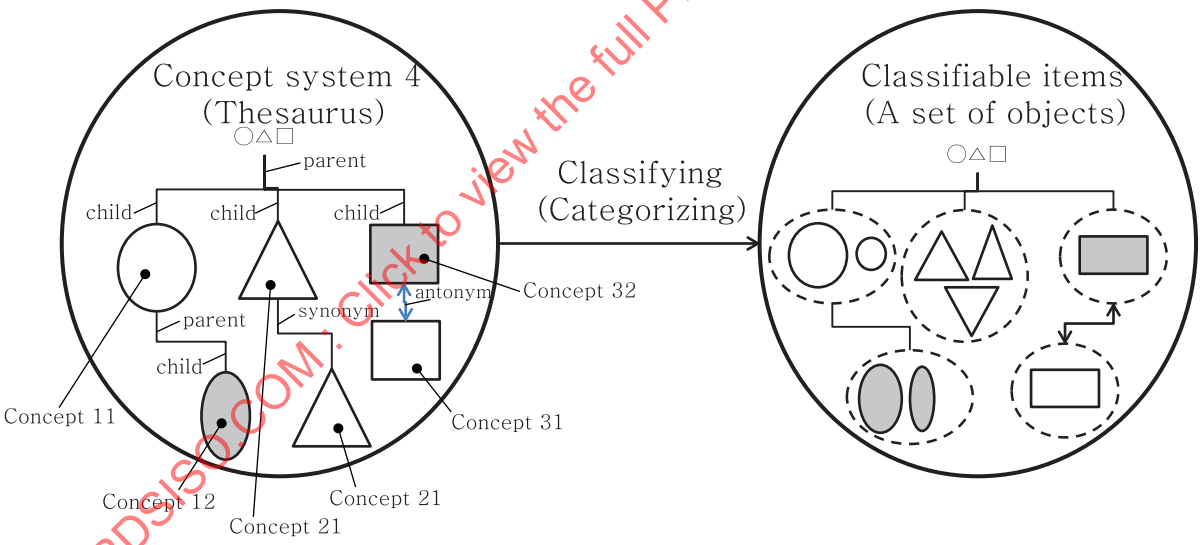

Figure 5 — Concept of classification (thesaurus)

Ontology was defined as an explicit specification of a conceptualization[4] (see Figure 6). A formal ontology is a controlled vocabulary expressed in an ontology representation language. This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints[5].

Ontologies are divided into axiomatized ontologies and linguistic ontologies. Axiomatized ontologies represent an understanding of a part of the world related to an information system. They may include business rules, theories and constraints, and serve as formal specifications for automatically generated database schema and applications. Linguistic ontologies may support natural language interfaces for queries of structured and unstructured data, extraction of information from text and translation systems.

EXAMPLE WordNet, geopolitical ontology, Cyc, gene ontology, etc.

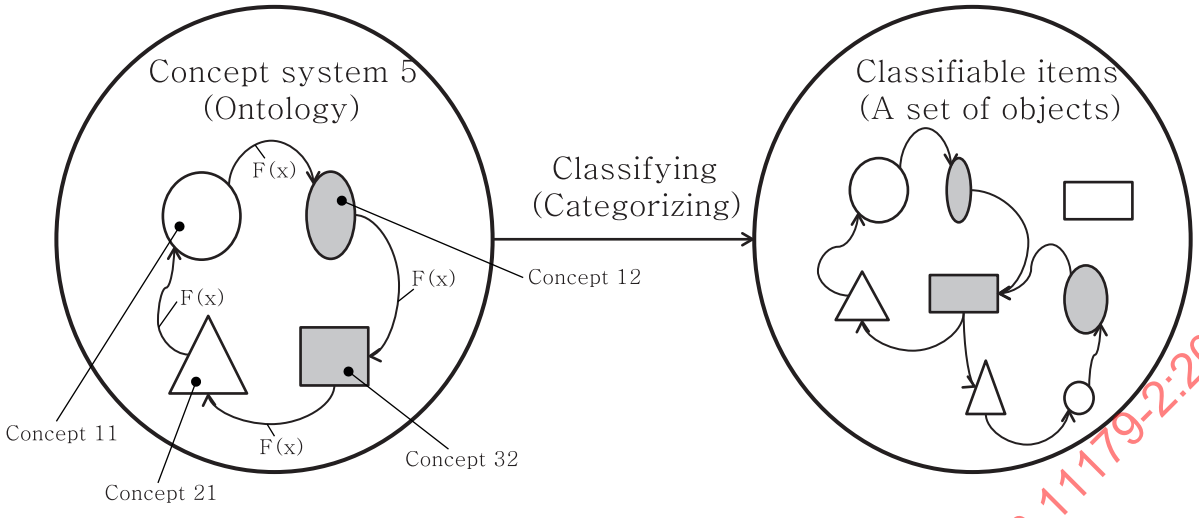

Figure 6 — Concept of classification (ontology)

7 Requirements for good classification scheme

A good classification scheme should satisfy several conditions such as uniqueness, consistency, comprehensiveness, continuity, extensibility and interoperability [6].

— **Uniqueness** – Every item of a classification scheme should be **unique** within the classification scheme. Duplication or overlapping items should not occurr within a classification scheme.

— **Consistency** – The classification scheme should be made **consistent** across all items of the classification scheme. The levels or meanings of items of the classification scheme should be balanced among them.

— **Comprehensiveness** – While a classification scheme created for a specific purpose within an MDR for its own internal use may be limited to a specific domain of interest, generally the scope of a classification scheme should **comprehensively** cover all ingredients of the domain suitable to the purpose of the classification scheme. There should be no part of the domain which is unclassified by the classification scheme.

— **Continuity** – When a classification scheme is updated, the new one should provide continuity with the old one such that the concepts and any identifiers for those concepts maintain the same meaning across versions.

— **Extensibility** – A classification scheme should be **extensible** to accommodate newly added items in the future.

— **Interoperability** – A classification scheme should be **interoperable** to other classification schemes in the same domain. Interoperability of classification scheme can be enhanced through the use of a common ontology, or top level ontology.