INTERNATIONAL STANDARD

ISO
12620

First edition
1999-10-01

Computer applications in terminology — Data categories

Aides informatiques en terminologie — Catégories de données

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard ISO 12620 was prepared by Technical Committee ISO/TC 37, *Terminology (principles and coordination)*, Subcommittee SC 3, *Computer applications.*

Annex A forms an integral part of this International Standard. Annexes B, C, D and E are for information only.

iv

Introduction

Terminological data are collected, managed, and stored in a wide variety of environments. For purposes of storage and retrieval, these data are organized into terminological entries, each of which traditionally treats information associated with a single concept. Data items appearing in individual terminological entries are themselves identified according to data category. Differences in approach and individual system objectives inevitably lead to variations in data category definition and in the assignment of data category names. The use of uniform data category names and definitions, at least at the interchange level, contributes to system coherence and enhances the reusability of data.

For terminology work in general, the following International Standards are relevant: ISO 704, ISO 860, ISO 1087, ISO 10241.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Computer applications in terminology — Data categories

1 Scope

This International Standard specifies data categories for recording terminological information in both computerized and non-computerized environments and for the interchange and retrieval of terminological information independent of the local software applications or hardware environments in which these data categories are used.

Some of the data categories specified in this International Standard derive requirements from definitions taken from other ISO terminology standards, e.g., ISO 1087 and ISO 5127, so as to harmonize the content of data categories across systems and to facilitate data interchange. The systematic arrangement of data categories in this International Standard is determined by database management considerations and does not reflect any theoretical arrangement of related terms and concepts used in other standards.

If applied for the purpose of interchanging machine-readable terminology, it is recommended that this International Standard be used in conjunction with ISO 12200, although it can also be used for modeling terminological information independent of computer applications.

It is not the purpose of this International Standard to specify actual data categories used in local database applications, because translation routines can be employed to convert application data categories to the universal categories specified here, provided that the values of the data associated with these categories are harmonized according to the relevant data category definitions.

It is also not the purpose of this International Standard to specify precise relationships among data categories, such as repeatability and combinability. These features remain the province of the individual database application and of ISO 12200. Nor does this International Standard prescribe a mandatory base set of data categories for local applications.

This International Standard does not specify data categories for the markup of bibliographic data entries. ISO 12083:1994, annex B, includes bibliographic elements among other elements cited as useful in the preparation and markup of machine-readable documents. Annex B of this International Standard lists data categories that correspond to elements cited in ISO 12083 and that are appropriate for the documentation of bibliographic citations in terminological entries.

> NOTE – ISO 12200:1999, annex B, provides additional information on the application of the bibliographic elements in ISO 12083 to the documentation of bibliographic information in terminological entries for those cases involving exchange of machine-readable terminology.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent

editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 639:1988, *Code for the representation of names of languages.*

ISO 639-2:1998, *Code for the representation of names of languages — Part 2: Alpha-3 code.*

ISO 1087:1990, *Terminology – Vocabulary.*

ISO 1087-2:-[1], *Terminology – Vocabulary – Part 2: Computer applications*

ISO 3166-1: 1997, *Codes for the representation of names of countries and their subdivisions – Part 1: Country codes.*

ISO 3166-2:1998, *Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code.*

ISO 3166-3:1999, *Code for the representation of names of countries and their subdivisions – Part 3: Codes for formerly used names of countries.*

ISO 5127-2:1983, *Documentation and information – Vocabulary – Part 2: Traditional documents.*

ISO 5127-6:1983, *Documentation and information – Vocabulary – Part 6: Documentary languages.*

ISO 8601:1988, *Data elements and interchange formats – Information interchange — Representation of dates and times.*

ISO 12200:1999, *Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange.*

ISO 12083:1994, *Information and documentation – Electronic manuscript preparation and markup.*

3 Terms and definitions

For the purpose of this International Standard, the terms and definitions given in ISO 1087-2 apply.

4 Selection of data categories

The categories employed in any given environment shall be carefully selected to meet the needs of those who create and use the terminology resource in question. For this purpose, an appropriate subset of the data categories specified in annex A shall be selected for application in the individual system.

---

1)To be published.

2

It is recommended that designers of terminology databases and other collections ensure that the content of data categories used in their systems conform to the content defined in these data category specifications. Terminological data prepared for interchange shall conform to the data category names and descriptions specified in annex A.

5   Documentation of data categories

The documentation of a data category for any terminological data collection shall enable the user to unambiguously interpret the content of the data category. The specification shall permit the user to differentiate the content of the data category from the content of other data categories used within the terminological data collection and the data categories used by interchange partners.

The standard specifications of data categories in annex A are based on standardized concepts for terminology management defined in ISO 1087 and ISO 5127-2 and 5127-6. If a data category required in a local application is not found in annex A, it shall be based on a well-established and documented concept in the field treated by a specific application. In any case, system designers shall notify the agency listed in annex E whenever they intend to create new data categories for use with the interchange format specified in ISO 12200.

If codes or representations other than those used in natural languages are used in conjunction with a data category, appropriate information for unambiguous interpretation shall be included in the specification of that data category. The specifications in annex A provide coding options wherever standardized codes are available.

6   Standard specification of data categories for use in terminology management

6.1   Format of the data category entry

The data category specifications included in annex A conform to the following format. (Not all categories appear in every specification.)

Specification category	Representation
Notation number:	**boldface number**
Preferred data category name:	**boldface**
Admitted name:	ADMITTED NAME: **boldface** [repeatable]
Full form:	FULL FORM: **boldface**
Related name:	RELATED NAME: **boldface** [repeatable]
Nonadmitted name:	NONADMITTED NAME: **boldface** [repeatable]
Data category description:	DESCRIPTION:
Note:	NOTE: [repeatable]
Permissible instances:	PERMISSIBLE INSTANCES: *italics*
Example:	EXAMPLE: [repeatable]

6.2 Typology of data categories

The data category specifications in annex A are divided into three major groups: data categories for terms and term-related information, descriptive data, and administrative data. The groups are further subdivided into ten sub-groups.

Term and term-related data categories:

Subgroup 1 consists of the data category term and contains a term or other information treated as if it were a term (e.g., phraseological units and standard text).

Subgroup 2 specifies data categories for term-related information.

Subgroup 3 specifies data categories for information relating to equivalence between or among terms assigned to the same or very similar concepts.

Descriptive data categories:

Subgroup 4 specifies data categories for the classification of concepts into subject fields and subfields, along with other classification-related information.

Subgroup 5 specifies data categories for concept-related description, i.e., different kinds of definitions, explanations and contextual material provided to define or otherwise determine the subject field and concept to which a term is assigned.

Subgroup 6 specifies data categories for indicating relations between pairs of concepts.

Subgroup 7 specifies data categories used to express the position of concepts within concept systems.

Subgroup 8 specifies the data category *note*. This category stands alone because it can be associated with any one of the other categories and therefore cannot be subordinated to any other specific subgroup.

Administrative data categories:

Subgroup 9 specifies data categories for documentary languages and thesauri.

Subgroup 10 specifies data categories for all other strictly administrative information.

Annex A

(normative)

Data categories

A.1   term

DESCRIPTION: A designation of a defined concept in a special language by a linguistic expression.

NOTE: For definition of related term, see ISO 1087:1990, 5.3.1.2

EXAMPLE: "radix" in annex C, figure C.1.

NOTE: Terms can consist of single words or be composed of multiword strings. The distinguishing characteristic of a term is that it is assigned to a single concept, as opposed to a phraseological unit, which combines more than one concept in a lexicalized fashion to express complex situations. *Quality assurance system* is a term, whereas *satisfy quality require-ments* is a phraseological unit, specifically a collocation.

A.2   term-related information

A.2.1 term type

DESCRIPTION: An attribute assigned to a term.

NOTE: *Term types* can include:

A.2.1.1   main entry term

ADMITTED NAME: **head term**

DESCRIPTION: The concept designation that has been chosen to head a terminological record.

EXAMPLE: "radix" in annex C, figure C.1

A.2.1.2   synonym

DESCRIPTION: Any term that represents the same or a very similar concept as the main entry term in a term entry.

EXAMPLE: "dynamicizer" in annex C, figure C.2

NOTE: Synonymy is generally relative, i.e., synonyms rarely cover all aspects of the same concept in all instances. The resulting *degree of synonymy* (A.2.10) is treated using the conventions defined for *degree of equivalence* (see A.3.1)

A.2.1.3   quasi-synonym

ADMITTED NAME: **near synonym**

DESCRIPTION: A term that represents the same or a very similar concept as another term in the same language, but for which interchangeability is limited to some contexts and inapplicable in others.

EXAMPLE: The distinction between synonyms and quasi-synonyms can be subjective or strongly context-oriented. For instance, some can designate "domain" and "subject field" as synonyms, whereas others would argue that "domain" is broader, but is usable in many of the same contexts and is therefore a quasi-synonym.

A.2.1.4   international scientific term

DESCRIPTION: A term that is part of an international scientific nomenclature as adopted by an appropriate scientific body.

EXAMPLE: *Homo sapiens*

A.2.1.5   common name

DESCRIPTION: A synonym for an international scientific term that is used in general discourse in a given language.

NOTE: Common names are generally formed based on metaphor, analogy, and function without reference to the classification rules applied to scientific nomenclatures. Although common names are widely used in general language, they are used in technical and scientific writing as well. Common names vary from language to language and frequently regionally within languages.

EXAMPLE: *Kalmia latifolia* is commonly called "mountain laurel" in the northern United States, and "calico bush" or "sheep's bane" in the south.

A.2.1.6   internationalism

DESCRIPTION: A term that has the same or nearly identical orthographic or phonemic form in many languages.

NOTE: Internationalisms frequently reflect Latin, Greek or English origins, but other languages, such as Arabic, French, Russian, Chinese and Japanese, have also contributed to the creation of internationalisms.

EXAMPLE: *en* alcohol, *fr* alcool, *de* Alkohol [from Arabic *al·kuhl*]

A.2.1.7   full form

ADMITTED NAME1: **expanded form**

ADMITTED NAME2: **expansion**

DESCRIPTION: The complete representation of a term for which there is an abbreviated form.

EXAMPLE: See the examples of *full forms* in A.2.1.8, ff.

A.2.1.8   abbreviated form of term

ADMITTED NAME: **abbreviated form**

DESCRIPTION: A term resulting from the omission of any part of the full term while designating the same concept.

NOTE 1: For definition of related term, see ISO 1087:1990, 5.5.2.

EXAMPLE: See individual abbreviation types listed below.

NOTE 2: Types of *abbreviated form* can include:

> abbreviation
> short form
> initialism
> acronym
> clipped term

NOTE 3: Each abbreviated form is derived from the full form of the term.

A.2.1.8.1 abbreviation

DESCRIPTION: An abbreviated form of a simple term resulting from the omission of some of its letters.

NOTE: For definition of related term, see ISO 1087:1990, 5.5.2.1.

EXAMPLE: full form: adjective
               abbreviation: adj.

A.2.1.8.2 short form of term

ADMITTED NAME: **short form**

DESCRIPTION: A variant of a multiword term that includes fewer words than the full form of the term.

EXAMPLES: full form: Intergovernmental Group of Twenty-four on International Monetary Affairs
              short form: Group of Twenty-four

NOTE: Many short forms are associated with long proper nouns, such as the names of governmental agencies, chemical compounds, and the like.

A.2.1.8.3 initialism

DESCRIPTION: An abbreviated form of a term consisting of some of the initial letters of the words making up a multiword term or the term elements making up a compound term when these letters are pronounced individually.

EXAMPLE: full form: bovine spongiform encephalopathy
         initialism: BSE

NOTE: The distinction between acronyms and initialisms can vary from language to language. The description given here applies to English.

A.2.1.8.4 acronym

DESCRIPTION: An abbreviated form of a term made up of letters from the full form of a multiword term strung together into a sequence pronounced only syllabically.

NOTE 1: An acronym can be so widely accepted that it becomes a term in its own right (e.g., *radar* in the following example).

NOTE 2: For definition of related term, see ISO 1087:1990, 5.5.2.1.2.

EXAMPLE: radar = radio detecting and ranging

A.2.1.8.5 clipped term

ADMITTED NAME: **truncated term**

DESCRIPTION: An abbreviated form of a term resulting from the omission of one or more term elements or syllables.

NOTE: For definition of related term, see ISO 1087:1990, 5.7.1.

EXAMPLE: full form: influenza
         clipped term: flu

A.2.1.9   variant

DESCRIPTION: One of the alternate forms of a term.

EXAMPLE: spelling variants: catalogue (GB), catalog (US)

A.2.1.10 transliterated form

DESCRIPTION: A form of a term resulting from an operation whereby the characters of an alphabetic writing system are represented by characters from another alphabetic writing system.

EXAMPLE: Cyrillic script: окружающая среда
                  Latin script: okružaûŝaâ sreda (environment)
                                      (Transliteration according to ISO 9:1995)
                                      okruzhayushchaya sreda
                                      (Transliteration according to BS 2979:1958)

A.2.1.11 transcribed form

DESCRIPTION: A form of a term resulting from an operation whereby the characters of one writing system are represented by characters from another writing system, taking into account the pronunciation of the characters converted.

EXAMPLE: Japanese:

Hiragana syllabary transcription:

Romanization according to ISO 3602:1989:   taihuu

English transcription: typhoon

Russian transcription: тайфун

Definition: a tropical storm in the western area of the Pacific Ocean in late summer and autumn

Chinese:

Romanization according to ISO 7098:1991:   gongfu

English transcription: kung fu

Russian transcription: кун фу

Definition: one of the Chinese martial arts

A.2.1.12 romanized form

DESCRIPTION: A form of a term resulting from an operation whereby non-Latin writing systems are converted to the Latin alphabet.

NOTE: Romanization is a specific form of transcription.

EXAMPLE: See example in A.2.1.10 and A.2.1.11.

A.2.1.13 symbol

DESCRIPTION: A designation of a concept by letters, numerals, pictograms or any combination thereof.

NOTE: For definition of related term, see ISO 1087:1990, 5.3.1.1

EXAMPLE: The symbol § can be used to represent a clause or subclause in a legal document.

A.2.1.14 formula

DESCRIPTION: Figures, symbols or the like used to express a concept briefly, such as a mathematical or chemical formula.

NOTE: A formula can function as a *term* representing the concept. In some cases, no other representation exists.

EXAMPLE: H_2O is the chemical formula for *water*.

A.2.1.15 equation

DESCRIPTION: An expression used to represent a concept based on a statement that two mathematical expressions are, for instance, equal as identified by the equal sign ($=$), or assigned to one another by a similar sign.

EXAMPLE: $E = mc^2$

NOTE: Such statements are sometimes documented in terminology databases.

A.2.1.16 logical expression

DESCRIPTION: An expression used to represent a concept based on mathematical or logical relations, such as statements of inequality, set relationships, boolean operations, and the like.
EXAMPLE: $x \neq y$, $x \in y$, x NOT y , etc.

A.2.1.17 materials management categories

A.2.1.17.1 sku
FULLFORM: **stockkeeping unit**
DESCRIPTION: An inventory item identified by a unique alphanumeric designation assigned to an object in an inventory control system.
EXAMPLE: For the catalog entry: "PLAID FLANNEL PANTS #5193 Sizes 3, 4, 6, 7, 10, 12", "#5193-6" represents a *sku* for the item: Style number #5193, size 6.
NOTE: Terminology databases that are linked to inventory control systems and manufacturing logistical systems include *skus* and *part numbers,* which act as designations within the system representing the object in question. Hence they function much like terms and even take on the character of terms in common discourse and text creation.

A.2.1.17.2 part number
DESCRIPTION: A unique alphanumeric designation assigned to an object in a manufacturing system.
EXAMPLE: Sample part numbers from an automotive power train manufacturing system, where each segment of the number represents a different classification level within the system:

clutch cover   1 110 036 00 a
driven disk flange   3 125 125 04 b
driven disk retainer plate   3 124 119 01 a
driven disk cover plate   3 122 234 00 c
diaphragm spring   4 220 100 00 g

NOTE: Terminology databases that are linked to inventory control systems and manufacturing logistical systems include *skus* and *part numbers*, which function as designations within the system representing the object in question. Hence they function much like terms and even take on the character of terms in common discourse and text creation.

A.2.1.18 phraseological unit
DESCRIPTION: Any group of two or more words that form a unit, the meaning of which frequently cannot be deduced based on the combined sense of the words making up the phrase.
NOTE: Although they are made up of more than one word and frequently contain more than one concept, phraseological units can be treated as individual terminological units in terminology databases. In this sense they are grouped together with "terms". They can, however, also be treated as contextual material in some databases.
EXAMPLES: See examples in A.2.1.18.1-A.2.1.18.3.