SLOVENSKI STANDARD
SIST ISO 12620:2019

01-oktober-2019

Upravljanje terminoloških virov - Specifikacije za podatkovne kategorije

Management of terminology resources -- Data category specifications

Terminologie et autres ressources langagières et ressources de contenu -- Spécifications des catégories de données

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Ta slovenski standard je istoveten z: ISO 12620:2019

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

ICS:

01.020	Terminologija (načela in koordinacija)	Terminology (principles and coordination)
35.240.30	Uporabniške rešitve IT v informatiki, dokumentiranju in založništvu	IT applications in information, documentation and publishing

SIST ISO 12620:2019 **en,fr,de**

2003-01.Slovenski inštitut za standardizacijo. Razmnoževanje celote ali delov tega standarda ni dovoljeno.

SIST ISO 12620:2019

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

SIST ISO 12620:2019

INTERNATIONAL STANDARD

ISO 12620

Third edition
2019-05

Management of terminology resources — Data category specifications

Gestion des ressources terminologiques — Spécifications des catégories de données

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

Reference number
ISO 12620:2019(E)

© ISO 2019

SIST ISO 12620:2019

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

COPYRIGHT PROTECTED DOCUMENT

© ISO 2019

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

SIST ISO 12620:2019

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)
SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

© ISO 2019 – All rights reserved

SIST ISO 12620:2019

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.

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 ISO documents 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 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).

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

For an explanation on 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 the following URL: www.iso.org/iso/foreword.html.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 3, *Management of terminology resources*.

This third edition cancels and replaces the second edition (ISO 12620:2009), which has been technically revised.

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb8aaa/sist-iso-12620-2019

The main changes compared to the previous edition are as follows.

ISO 12620:2009, *Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources*, described a data model and management features for a Data Category Registry designed for the purpose of standardizing data category specifications. The current edition of ISO 12620 has been streamlined to eliminate the standardization function previously built into the data model. It describes requirements for maintaining a consensus-based, industry-appropriate repository of harmonized data category specifications for use in language resources.

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.

© ISO 2019 – All rights reserved

SIST ISO 12620:2019

Introduction

Data associated with language resources are identified, collected, managed and stored in a wide variety of environments. Data appearing in language resources are generalized into classes that are referred to as *data categories*. Differences in approach for developing different kinds of language resources as well as differences in technical environments inevitably lead to variations in data category definitions and data category names. The use of uniform data category names and definitions employed in resources within the same linguistic domain (for example, among terminological resources, lexicographical resources, annotated text corpora, etc.) contributes to system coherence and enhances the re-usability of data. Such uniform use requires access to formal data category specifications. Defining a clear framework for specifying, managing and using data categories will increase interoperability of language resources.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

SIST ISO 12620:2019

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

SIST ISO 12620:2019

INTERNATIONAL STANDARD **ISO 12620:2019(E)**

Management of terminology resources — Data category specifications

1 Scope

This document provides guidelines and requirements governing data category specifications for language resources. It specifies mechanisms for creating, documenting, harmonizing and maintaining data category specifications in a data category repository. It also describes the structure and content of data category specifications. The intended audience of this document is researchers and practitioners in fields of language resource management who use data categories and data category specifications.

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 24619, *Language resource management — Persistent identification and sustainable access (PISA)*

iTeh STANDARD PREVIEW
(standards.iteh.ai)

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:

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01e63aaa/sist-iso-12620-2019

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
conceptual domain
permissible content of a *data category* (3.2)

EXAMPLE In a terminology database, the data category /part of speech/ could have a conceptual domain consisting of the values: /noun/, /verb/, /adjective/, /adverb/.

Note 1 to entry: The permissible content can be enumerated (such as in a pick list), as in the example, or subject to formal restrictions such as dates, or free text such as the conceptual domain of /definition/. Although the latter type is not formally restricted, it is nevertheless subject to adherence to the requirements of its data category specification, i.e., it contains a true definition and not a note, example, or some other piece of information.

3.1.1
open conceptual domain
conceptual domain (3.1) that has no formal restrictions

Note 1 to entry: An open conceptual domain is frequently associated with data categories that take free text as their content, such as /definition/.

Note 2 to entry: Some requirements are not machine processable, for instance, to require that /definition/ only contain definitional information.

3.1.2
closed conceptual domain
conceptual domain (3.1) that is restricted to a set of enumerated values

EXAMPLE The data category /grammatical gender/ can have a conceptual domain consisting of the values

SIST ISO 12620:2019

/feminine/, /masculine/ and /neuter/.

3.1.3
constrained conceptual domain
conceptual domain (3.1) that is restricted to a constraint or rule specified in a schema-specific language

EXAMPLE The data category /date/ can be constrained by a system setting to certain date formats, or a data category can be subject to a termbase-specific rule, such as making it mandatory to enter a /source/ for a /definition/.

3.1.4
simple conceptual domain
conceptual domain (3.1) that has only two values

Note 1 to entry: The two values can be "yes" or "no", "true" or "false", or other such binary representation.

3.2
data category
DC
class of data items that are closely related from a formal or semantic point of view

EXAMPLE /part of speech/, /subject field/, /definition/.

Note 1 to entry: A data category can be viewed as a generalization of the notion of a field in a database.

Note 2 to entry: In running text, such as in this document, data category names are enclosed in forward slashes (e.g. /part of speech/).

[SOURCE: ISO 30042:2019, 3.8]

iTeh STANDARD PREVIEW
(standards.iteh.ai)
SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

3.2.1
open data category
data category (3.2) that has an *open conceptual domain* (3.1.1)

3.2.2
closed data category
data category (3.2) that has a *closed conceptual domain* (3.1.2)

3.2.3
constrained data category
data category (3.2) that has a *constrained conceptual domain* (3.1.3)

3.2.4
simple data category
data category (3.2) that has a *simple conceptual domain* (3.1.4)

Note 1 to entry: See also *pick list value* (3.9).

3.3
data category concept
semantic content of a *data category* (3.2), independent of any specific implementations

3.4
data category name
linguistic representation of a *data category* (3.2) as it appears in a particular language or in a particular application or language resource

EXAMPLE The data category name for /part of speech/ is "part of speech" in English, and "partie du discours" in French.

© ISO 2019 – All rights reserved

SIST ISO 12620:2019

3.5
data category repository
DCR
digital repository of *data category specifications* (3.7)

Note 1 to entry: Data category repositories are used as references when specifying language resources.

Note 2 to entry: A DCR for language resources is available at www.datcatinfo.net.

3.6
data category selection
DC selection
set of *data category specifications* (3.7) selected from a *data category repository* (3.5)

Note 1 to entry: A data category selection can represent the *data categories* (3.2) used within a research discipline or a specific application or project.

3.7
data category specification
DC specification
complete descriptive record of a *data category* (3.2)

3.8
persistent identifier
PID
unique Uniform Resource Identifier (URI) that provides permanent access to a digital object independently of its physical location or current ownership

iTeh STANDARD PREVIEW

EXAMPLE http://www.datcatinfo.net/datcat/DC-70.

(standards.iteh.ai)

[SOURCE: ISO 24619:2011, 3.2.4, modified — order of terms inverted, definition slightly reworded, note deleted, example added.]

SIST ISO 12620:2019
https://standards.iteh.ai/catalog/standards/sist/46beadf9-4413-4dda-b806-0d5d01eb3aaa/sist-iso-12620-2019

3.9
pick list value
one of the enumerated or permissible values of a *closed data category* (3.2.2)

EXAMPLE "singular" and "plural" are pick list values in a field labelled "Grammatical Number".

Note 1 to entry: See also *simple data category* (3.2.4).

Note 2 to entry: Due to data modelling variance, most types of information that can be represented as pick list values in a database can also be represented as simple data categories. For example, "Plural" can be implemented as a checkbox, which, when checked, takes the value "yes" and when unchecked, takes the value "no".

4 Data categories and data category specifications

A data category (DC) is a class of information that forms part of a data collection or annotation scheme for a given language resource. For example, /definition/ and /part of speech/ are common data categories in terminological and lexicographical resources. Data category names can appear as the name of a field in the user interface of a software application, or as a markup element in an annotated resource.

Some data categories are pertinent to a specific application, research area, or type of resource and not others. For instance, a /concept identifier/ is characteristic of terminological or ontological resources, whereas /sense number/ is applicable to lexicographical resources. On the other hand, many data categories, frequently those of a strictly linguistic nature such as /part of speech/, /grammatical gender/ and /grammatical number/, are common to a wide variety of resources. These data categories may not always be implemented in the same way in different resources or applications, but each nevertheless evokes one universal data category concept. For instance, for terminology management, only a small