

SVENSK STANDARD

SS-ISO 16642:2018



Fastställt/Approved: 2018-04-18
Publicerad/Published: 2018-04-18
Utgåva/Edition: 1
Språk/Language: engelska/English
ICS: 01.020; 35.240.30

Datortillämpningar inom terminologi – Ramverk för terminologisk märkning (ISO16642:2017, IDT)

Computer applications in terminology – Terminological markup framework (ISO16642:2017, IDT)

This preview is downloaded from www.sis.se. Buy the entire standard via <https://www.sis.se/std-80003634>

Standarder får världen att fungera

SIS (Swedish Standards Institute) är en fristående ideell förening med medlemmar från både privat och offentlig sektor. Vi är en del av det europeiska och globala nätverk som utarbetar internationella standarder. Standarder är dokumenterad kunskap utvecklad av framstående aktörer inom industri, näringsliv och samhälle och befrämjar handel över gränser, bidrar till att processer och produkter blir säkrare samt effektiviserar din verksamhet.

Delta och påverka

Som medlem i SIS har du möjlighet att påverka framtida standarder inom ditt område på nationell, europeisk och global nivå. Du får samtidigt tillgång till tidig information om utvecklingen inom din bransch.

Ta del av det färdiga arbetet

Vi erbjuder våra kunder allt som rör standarder och deras tillämpning. Hos oss kan du köpa alla publikationer du behöver – allt från enskilda standarder, tekniska rapporter och standardpaket till handböcker och onlinetjänster. Genom vår webbtjänst e-nav får du tillgång till ett lättnavigerat bibliotek där alla standarder som är aktuella för ditt företag finns tillgängliga. Standarder och handböcker är källor till kunskap. Vi säljer dem.

Utveckla din kompetens och lyckas bättre i ditt arbete

Hos SIS kan du gå öppna eller företagsinterna utbildningar kring innehåll och tillämpning av standarder. Genom vår närhet till den internationella utvecklingen och ISO får du rätt kunskap i rätt tid, direkt från källan. Med vår kunskap om standarders möjligheter hjälper vi våra kunder att skapa verklig nytta och lönsamhet i sina verksamheter.

Vill du veta mer om SIS eller hur standarder kan effektivisera din verksamhet är du välkommen in på www.sis.se eller ta kontakt med oss på tel 08-555 523 00.



Standards make the world go round

SIS (Swedish Standards Institute) is an independent non-profit organisation with members from both the private and public sectors. We are part of the European and global network that draws up international standards. Standards consist of documented knowledge developed by prominent actors within the industry, business world and society. They promote cross-border trade, they help to make processes and products safer and they streamline your organisation.

Take part and have influence

As a member of SIS you will have the possibility to participate in standardization activities on national, European and global level. The membership in SIS will give you the opportunity to influence future standards and gain access to early stage information about developments within your field.

Get to know the finished work

We offer our customers everything in connection with standards and their application. You can purchase all the publications you need from us - everything from individual standards, technical reports and standard packages through to manuals and online services. Our web service e-nav gives you access to an easy-to-navigate library where all standards that are relevant to your company are available. Standards and manuals are sources of knowledge. We sell them.

Increase understanding and improve perception

With SIS you can undergo either shared or in-house training in the content and application of standards. Thanks to our proximity to international development and ISO you receive the right knowledge at the right time, direct from the source. With our knowledge about the potential of standards, we assist our customers in creating tangible benefit and profitability in their organisations.

If you want to know more about SIS, or how standards can streamline your organisation, please visit www.sis.se or contact us on phone +46 (0)8-555 523 00



Den internationella standarden ISO 16642:2017 gäller som svensk standard. Detta dokument innehåller den officiella engelska versionen av ISO 16642:2017.

The International Standard ISO 16642:2017 has the status of a Swedish Standard. This document contains the official version of ISO 16642:2017.

© Copyright/Upphovsrätten till denna produkt tillhör SIS, Swedish Standards Institute, Stockholm, Sverige. Användningen av denna produkt regleras av slutanvändarlicensen som återfinns i denna produkt, se standardens sista sidor.

© Copyright SIS, Swedish Standards Institute, Stockholm, Sweden. All rights reserved. The use of this product is governed by the end-user licence for this product. You will find the licence in the end of this document.

Uppllysningar om sakinnehållet i standarden lämnas av SIS, Swedish Standards Institute, telefon 08-555 520 00. Standarder kan beställas hos SIS som även lämnar allmänna upplysningar om svensk och utländsk standard.

Information about the content of the standard is available from the Swedish Standards Institute (SIS), telephone +46 8 555 520 00. Standards may be ordered from SIS, who can also provide general information about Swedish and foreign standards.

Denna standard är framtagen av kommittén för Språk och terminologi, SIS/TK 115.

Har du synpunkter på innehållet i den här standarden, vill du delta i ett kommande revideringsarbete eller vara med och ta fram andra standarder inom området? Gå in på www.sis.se - där hittar du mer information.

Contents

Page

Foreword	iv
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Modular approach	4
5 Generic model for describing terminological data	5
5.1 Principles.....	5
5.2 Generic representation of components and information units.....	6
5.3 The metamodel.....	8
5.4 Example.....	10
6 Requirements for compliance to TMF	11
7 Interchange and interoperability	12
8 Representing languages	12
9 Defining a TML	13
9.1 Steps.....	13
9.2 Defining interoperability conditions.....	13
10 Implementing a TML	13
10.1 General.....	13
10.2 Implementing the metamodel.....	13
10.3 Anchoring data categories on the XML outline.....	14
10.3.1 General.....	14
10.3.2 Styles and vocabulary.....	14
10.4 Constraints on datatypes.....	15
10.5 Implementing annotations.....	15
10.6 Implementing brackets.....	15
Annex A (informative) Conformance of terminological data to TMF: example scenario	16
Bibliography	21

SS-ISO 16642:2018 (E)

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.

This document was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 3, *Computer applications for terminology*.

This second edition cancels and replaces the first edition (ISO 16642:2003), which has been technically revised.

The main changes compared to the previous version are as follows:

- The following formats are no longer actively used. Consequently, references to these formats have been removed (including Annex A, Annex B, and Annex C):
 - Martif with specified constraints (MSC);
 - Geneter;
 - Data category interchange format (DCIF);
 - Generic mapping tool (GMT).
- With the removal of Annex B and Annex C, this document no longer includes any comprehensive code examples of a TML. Examples of TMLs are now available in ISO 30042, TermBase eXchange, and also at the following Web site: www.tbxinfo.net.
- References to the former ISO/TC 37 Data Category Registry or ISocat have been changed from normative to informative. In addition, the name has changed to DatCatInfo, now as an example of data category repositories.
- References to ISO 12620:1999 and ISO 12620:2009 have been removed. These previous standards have been withdrawn.
- The TypedValuedElement style has been added.
- Examples have been updated to reflect ISO 30042:2008 (TBX). TBX-Basic is mentioned as a TML.

- Some of the examples and tables have been moved to appropriate sections.
- As a consequence of the aforementioned changes, some historical, didactic, or duplicate information has been removed to adhere more closely to ISO editorial standards.

SS-ISO 16642:2018 (E)

Introduction

Terminological data are collected, managed and stored in a wide variety of systems, typically various kinds of database management systems, ranging from personal computer applications for individual users to large terminological database systems operated by major companies and governmental agencies. Terminology databases are comprised of various types of information, called data categories, and can adopt different structural models. However, terminological data often need to be shared and reused in a number of applications, and this sharing is facilitated when the data adheres to a common model. To facilitate co-operation and to prevent duplicate work, it is important to develop standards and guidelines for creating and using terminological data collections (TDCs) as well as for sharing and exchanging data.

This document presents a modular approach for analysing existing TDCs and designing new ones. It also provides a framework for defining terminological markup languages (TMLs) that are interoperable.

This document makes reference to DatCatInfo, an example of an available data category repository. DatCatInfo is an online database of information about the types of data that can be included in terminological data collections and other language resources. It is available at www.datcatinfo.net.

Computer applications in terminology — Terminological markup framework

1 Scope

This document specifies a framework for representing data recorded in terminological data collections (TDCs). This framework includes a metamodel and methods for describing specific terminological markup languages (TMLs) expressed in XML. The mechanisms for implementing constraints in a TML are defined, but not the specific constraints for individual TMLs.

This document is designed to support the development and use of computer applications for terminological data and the exchange of such data between different applications. This document also defines the conditions that allow the data expressed in one TML to be mapped onto another TML.

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 704, *Terminology work — Principles and methods*

ISO 1087-1, *Terminology work — Vocabulary — Part 1: Theory and application*

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

ISO 26162, *Systems to manage terminology, knowledge and content — Design, implementation and maintenance of terminology management systems*

ISO 30042:2008, *Systems to manage terminology, knowledge and content — TermBase eXchange (TBX)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087-1 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at <http://www.electropedia.org/>
- ISO Online browsing platform: available at <http://www.iso.org/obp>

3.1

basic information unit

information unit (3.12) attached to a *component* (3.3) of the metamodel and that can be expressed by means of a single *data category* (3.6)

3.2

complementary information

CI

information supplementary to that described in *terminological entries* (3.22) and shared across the *terminological data collection* (3.21)

Note 1 to entry: Domain hierarchies, institution descriptions, bibliographic references and references to text corpora are typical examples of complementary information.

SS-ISO 16642:2018 (E)

3.3 component
elementary description unit of a metamodel to which *data categories* (3.6) can be associated to form a data model

3.4 compound information unit
information unit (3.12) attached to a *component* (3.3) of the metamodel that is expressed by means of several grouped *data categories* (3.6), that, taken together, express a coherent unit of information

3.5 conceptual domain
set of valid value meanings associated with a *data category* (3.6)

Note 1 to entry: For example, the data category /part of speech/ could have the following conceptual domain: /noun/, /verb/, /adjective/, /adverb/, and so forth.

3.6 data category
elementary descriptor used in a linguistic description or annotation scheme

Note 1 to entry: In this document, data categories are indicated in between forward slashes (/), e.g. /definition/.

3.7 data category repository
DCR
electronic repository of *data category specifications* (3.9) to be used as a reference for the definition of linguistic annotation schemes or any other representation model for language resources

Note 1 to entry: A DCR for language resources is available at <http://www.datcatinfo.net>.

3.8 data category selection
DCS
set of *data categories* (3.6) selected from a *DCR* (3.7)

3.9 data category specification
set of attributes used to fully describe a given *data category* (3.6)

Note 1 to entry: The abbreviation “DCS” is associated with data category selection and is not used for data category specification.

3.10 expansion tree
structured group of XML elements that implement a level of the metamodel in a given *TML* (3.23)

3.11 global information
GI
technical and administrative information applying to the entire *terminological data collection* (3.21)

Note 1 to entry: For example, the title of the terminological data collection, revision history, owner or copyright information.

3.12 information unit
IU
elementary piece of information attached to a structural level of the metamodel

3.13
language section

LS

part of a *terminological entry* (3.22) containing information related to one language

Note 1 to entry: One terminological entry may contain information on one or more languages.

3.14
object language

language being described

3.15
persistent identifier

PID

unique Uniform Resource Identifier (URI) that assures permanent access for a digital object by providing access to it independently of its physical location or current ownership

3.16
structural node

instance of *component* (3.3) within the representation of a *terminological data collection* (3.21)

3.17
structural skeleton

abstract description of an instance of a *terminological data collection* (3.21) in conformity with the metamodel

3.18
style

specification for the implementation of a *data category* (3.6) in XML

3.19
term component section

TCS

part of a *term section* (3.20) giving linguistic information about the components of a term

3.20
term section

TS

part of a *language section* (3.13) giving information about a term

3.21
terminological data collection

TDC

resource consisting of *terminological entries* (3.22) with associated meta data and documentary information

3.22
terminological entry

TE

part of a *terminological data collection* (3.21) which contains the terminological data related to one concept

Note 1 to entry: Every element in the TE can be linked to complementary information, to other terminological entries and to other elements in the same terminological entry.

3.23
terminological markup language

TML

XML format for representing a *terminological data collection* (3.21) conforming to the constraints expressed in this document