
Computer applications in terminology — Terminological markup framework

*Applications informatiques en terminologie — Plate-forme pour le
balisage de terminologies informatisées*



PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

This document is a preview generated by EVS

© ISO 2003

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

	Page
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 General principles and interoperability principle	4
5 Generic model for describing linguistic data and its application to terminology	6
5.1 Introduction	6
5.1.1 General principles	6
5.1.2 Example	7
5.2 Generic representation of structural levels and information units	8
5.3 The terminological meta-model	9
5.4 Designing representations of terminological data on the basis of the meta-model	12
5.5 Interchange, dissemination and interoperability	12
5.6 XML canonical representation of the generic model	13
5.6.1 Introduction	13
5.6.2 Example	13
5.6.3 Description of the GMT format	14
5.7 Representing languages in a terminological data collection	17
6 Defining a TML	18
6.1 General	18
6.2 Defining interoperability conditions	18
6.3 Implementing a TML	18
6.3.1 Introduction	18
6.3.2 Implementing the meta-model	18
6.3.3 Anchoring data categories on the TML XML outline	19
6.3.4 Implementing annotations	20
6.3.5 Implementing brackets	21
6.3.6 Namespaces	21
Annex A (normative) XML schema of the GMT format	22
Annex B (normative) The MSC TML	24
Annex C (normative) The Geneter TML	29
Annex D (informative) Conformance of terminological data to TMF	43
Bibliography	48

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. 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.

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.

ISO 16642 was prepared by Technical Committee ISO/TC 37, *Terminology and other language resources*, Subcommittee SC 3, *Computer applications for terminology*.

Introduction

Terminological data are collected, managed and stored in a wide variety of systems, typically in applications, i.e. various kinds of database management system, ranging from personal computer applications for individual users to mainframe term-bank systems operated by major companies and governmental agencies. Termbases are comprised of various sets of data category and are based on various kinds of data model. Terminological data often need to be shared and reused in a number of applications, and this sharing is usually accomplished using intermediate formats. To facilitate co-operation and to prevent duplicate work, it is important to develop standards and guidelines for creating and using terminological data collections as well as for sharing and exchanging data.

The meta-model defined in this International Standard fits into an integrated approach to be used in analysing existing terminological data collections and in designing new ones, which are typically processed using relational or text-based data management systems. Terminological data collections can also be stored as structured documents with markup based on formats that are typically defined using Standard Generalized Markup Language (SGML), defined in ISO 8879 [12], or eXtensible Markup Language (XML), which is based on SGML but amended for use on the World Wide Web by the World Wide Web Consortium (W3C). An integrated approach eases the tasks of importing data from a flat file with markup into a database and of exporting data from a database to a structured document. Another motivation for an integrated approach, as opposed to entirely separate approaches for databases and structured documents, is that XML-based formats are now being processed in new ways, similar to traditional database management systems. For example, XML files are being queried and updated directly without importing data into traditional database environments.

This integrated approach to analysis and design consists of two levels of abstraction. The first (and most abstract) level of the integrated approach is the meta-model level. The meta-model level, which could also be called the abstract conceptual data model level, supports analysis and design at a very general level. The second level is the data model level.

At the data model level, the designer of the terminological data collection has the possibility to make various choices, based on real-life needs. First, designers must determine the form of representation most appropriate for their terminological data, addressing the following choices:

- whether to use a relational database or a flat file with markup;
- whether the data will be used primarily for queries and updates and be represented in some database management system and, if this is the case, what system to use;
- whether the data will be used primarily for sharing and interchange, and be represented in a flat file with markup.

For the purposes of this International Standard it is assumed that all flat files will use XML markup.

Once the choice between a database management system and a flat file with XML markup has been made, a data model must be chosen. For a relational database, a typical method of describing a data model is an entity-relationship diagram. For an XML document, a typical method of describing a data model is a Document Type Definition (DTD). An alternative method, using what is called an “XML schema”, is provided by the W3C. In the future, it will be possible to use more abstract methods of describing an XML format.

A specific implementation of the meta-model for terminology markup expressed in XML is called a terminological markup language (or TML), which can be described on the basis of a limited number of characteristics, namely

- how the TML expresses the structural organization of the meta-model (i.e. the expansion trees of the TML),
- the specific data categories used by the TML and how they relate to the meta-model,

- the way in which these data categories can be expressed in XML and thus anchored on the expansion trees of the TML, i.e. the XML style of any given data category, and
- the vocabularies used by the TML to express those various informational objects as XML elements and attributes according to the corresponding XML styles.

Some of the examples in this International Standard are instances of the MSC (MARTIF with Specified Constraints) and Geneter formats as described in Annex B and Annex C respectively.

This document is a preview generated by EVS

Computer applications in terminology — Terminological markup framework

1 Scope

This International Standard specifies a framework designed to provide guidance on the basic principles for representing data recorded in terminological data collections. This framework includes a meta-model and methods for describing specific terminological markup languages (TMLs) expressed in XML. The mechanisms for implementing constraints in a TML are defined in this International Standard, but not the specific constraints for individual TMLs, except for the three TMLs defined in Annexes B to D.

This International Standard is designed to support the development and use of computer applications for terminological data and the exchange of such data between different applications. It does not standardize data categories and methods for the specification of data structures which are specified in ISO 12620 and other related International Standards.

This International Standard also defines the conditions that allow the data expressed in one TML to be mapped onto another TML and specifies a generic mapping tool (GMT) for this purpose (see Annex A).

In addition, this International Standard describes a generic model for describing linguistic data.

2 Normative references

The following referenced documents are indispensable for the application 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 1087-1, *Terminology work — Vocabulary — Part 1: Theory and application*

ISO 1087-2, *Terminology work — Vocabulary — Part 2: Computer applications*

ISO 12620:1999, *Computer applications in terminology — Data categories*

Extensible Markup Language (XML) 1.0, Second edition, BRAY, T., PAOLI, J., SPERBERG-MCQUEEN, C. M., and MALER, E. (eds.), W3C Recommendation 6 October 2000, available at <<http://www.w3.org/TR/REC-xml>>

Dublin Core Qualifiers, 2000-07-11, available at <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>>

XHTMLTM 1.0 The Extensible HyperText Markup Language, 2nd edition, available at <<http://www.w3.org/TR/xhtml1/>>

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087-1 and ISO 1087-2, and the following apply.