Contents

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland
Internet iso@iso.ch

Printed in Switzerland

Figures, tables and examples cited in this standard

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 12200 was prepared by Technical Committee ISO/TC 37, *Terminology (principles and coordination)*, Subcommittee SC 3, *Computer applications.*

The specifications in this International Standard were developed in close cooperation with the Text Encoding Initiative (TEI) and the Localisation Industry Standards Association (LISA).

ISO 12200 is based on ISO 8879, Standard Generalized Markup Language (SGML). This International Standard covering negotiated interchange is designed to be as open and flexible as possible in order to cover all types and forms of terminological entry structures that occur in terminological databases and specialized dictionaries.

Further parts of ISO 12200 may specify more restricted interchange formats for specific purposes. The objective of these further parts would be to enable more data to be passed between systems without customized intervention. Such further parts of ISO 12200 would specify formats that will be backward compatible with this International Standard so that documents structured according to further parts of ISO 12200 would be parsable using the Document Type Definition (DTD) specified in this International Standard, but documents structured according to this International Standard would not necessarily be parsable with the DTD specified in one of the further parts of ISO 12200.

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

Introduction

Terminological data are collected, managed, and stored in a wide variety of terminology database systems, ranging from personal computer applications for individual users to mainframe term-bank systems operated by major companies and governmental agencies. The interchange of terminological data has become increasingly necessary among different applications, systems, and hardware platforms. This International Standard is designed to support these needs for efficient data interchange.

ISO 8879, which covers Standard Generalized Markup Language (SGML), provides a method of describing documents. Instead of encoding how a document is rendered on the page, it describes the structural properties of the document and the interrelation of the components making up the document. It is well-known that SGML provides a single universal descriptive language in which the many available markup systems can be represented to facilitate transfer of texts (i.e., of information) from one program or application to another. As the use of SGML grows, it is being more widely used, in accordance with the intentions of its designers, for marking up text for data interchange and information retrieval, as well as for encoding texts for manipulation in hypertext environments.

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

Computer applications in terminology — Machine-readable terminology interchange format (MARTIF) — Negotiated interchange

1 Scope

This International Standard is based on ISO 8879. It deals with negotiated interchange and is designed to be as open and flexible as possible in order to cover all types and forms of terminological entry structures that occur in terminological databases and specialized dictionaries, as well as among various applications, operating systems, and hardware platforms. ISO 12200 is primarily designed for use with terminological data that can be stored, read, retrieved and manipulated by a computer. It is not limited to any specific software or hardware configurations.

The primary purpose of this International Standard is to provide guidance for programmers and analysts in designing export and import software for data interchange between terminology databases. The Document Type Definition (DTD) specified in this International Standard permits partial validation of interchange files using a general-purpose SGML parser (i.e., confirmation that the document conforms to the structure specified by the DTD).

NOTE – Before an initial interchange between new partners, some level of data examination, negotiation, and adjustment of conversion routines can be necessary.

This International Standard can also be used for the creation of conversion routines to accommodate data encoded according to ISO 6156. It is recommended that this International Standard be used in conjunction with ISO 12620.

This International Standard does not specify the structure and function of individual databases.

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/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange.*

ISO 1087-2:—[1], *Terminology work — Vocabulary — Part 2: Computer applications.*

1) To be published.

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

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

ISO 8879:1986, *Information processing — Text and office systems — Standard Generalized Markup Language (SGML).*

ISO/IEC 10646-1:1993, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane.*

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

ISO 12620:1999, *Terminology — Computer applications — Data categories.*

3 Terms and definitions

For the purpose of this International Standard, the definitions given in ISO 8879 and ISO 1087-2 apply. For the convenience of users of this International Standard, some relevant definitions from ISO 8879 and ISO 1087-2 are contained in Annex F. The following definition was adapted to avoid ambiguity in the context of this International Standard.

3.1
attribute
<in MARTIF> characteristic quality of a **generic identifier**
NOTE — Adapted from ISO 8879:1986.

4 Structuring terminological information

The basic unit of terminological data management used in MARTIF documents shall be the terminological entry. In other words, a MARTIF document shall be made up of terminological entries. A terminological entry shall contain information pertaining to a specific concept or several closely related concepts, one or more terms in one or more languages, and other descriptive and administrative information deemed useful in a particular context.

NOTE — Terminological data can take the form of terminology databases or can be used to print hardcopy documents, technical and terminological dictionaries, vocabularies and - to a certain extent - documentation thesauri. For SGML applications, however, even terminology databases themselves can be viewed as documents. The structure and presentation of data vary considerably among terminology databases as a result of different user needs, approaches, and software requirements. These variations also reflect whether the entry is monolingual, bilingual, or multilingual, whether it contains prescriptive or descriptive information, and the work environment in which the terminology file is created and used.

In order to account for differences in database design, individual terminological entry structures shall be mapped to the MARTIF structure for interchange purposes. It shall be noted, however, that if the structure of the source database is richer than that of the target database, a potential loss of information can only be avoided by appropriately re-structuring and re-tagging the target database.

Figure 1 represents a typical terminological entry such as might be generated in a multilingual working environment.

Opazität

Maß für die Lichtundurchlässigkeit

Figure 1 ä— Sample terminological entry displayed by listing data categories and corresponding data category content

NOTE – This sample terminological entry represents a realistic working environment where information on a single concept has been taken from different sources in different languages and combined in a single terminological entry. Example 3 shows this terminological entry expressed as a MARTIF <termEntry>, and Example E.1 in Annex E incorporates the same <termEntry> into a full MARTIF document.

5 Terminological entries in MARTIF

5.1 Data categories

5.1.1 Specification of Data Categories

MARTIF is designed to allow interchange of terminological data residing in terminology databases of any structure. Therefore each data category within the terminological entry shall be properly identified and relationships among the data categories shall be encoded within the entry so that they can be redistributed to any required arrangement in the target database.

The generic identifiers (GIs or tag names) specified in 5.1.2 and attributes specified in 5.1.3 shall be used to mark up (i.e., to name) data categories when they occur in MARTIF documents. In addition, Annex A specifies the full normalized forms that shall be used for these data categories in the MARTIF environment, as well as the attribute values that shall be used with them (see 5.1.4).

Some of these data categories identify sub-categories of information related to terms and the concepts they represent. Others provide administrative information related to the terminological entry itself and to file management. The data categories listed in Annex A are defined in ISO 12620 and shall be used for encoding terminological data for interchange using MARTIF. For this purpose, data category names used in local applications that do not comply with ISO 12620 shall be converted accordingly. If a data category required in a local application is not available in ISO 12620, system designers should notify the coordinators of that standard accordingly (see ISO 12620, Annex E).

5.1.2 MARTIF tags

Table 1 lists the specific Generic Identifiers that shall be used in terminological entries and bibliographic references within the MARTIF environment.

The MARTIF DTD is designed around three primary categories represented by the generic identifiers <term>, <descrip>, and <admin>. Annex C lists these and other GIs together with the data categories associated with them as per the examples shown in Annexes A and B, as well as those data categories associated with specific attributes.

Table 1 MARTIF tags and their description

<termEntry>	Shall contain a single complete terminological entry for one concept expressed in one language and comprising one or more terms and their associated descriptive and administrative data, or, in bilingual and multilingual terminology work, two or more closely related concepts comprising one or more terms in each language and their associated descriptive and administrative data. Attributes include: *type*, which classifies the terminological entry as per the data categories specified in ISO 12620.

Table 1, continued

<langSet>	Language set; within a <termEntry> element, shall be used to group multiple <tig>s and <ntig>s associated with a single language. The attribute *lang* is required, unless inherited
<tig>	Terminological information group; within a <termEntry> element, shall contain information elements associated with a single term, all of which must function on the same level; i.e., embedding within the subordinate elements of the <tig> is not allowed. The attribute *lang* is required, unless inherited.
<ntig>	Nested terminological information group; shall be used within a <termEntry> when some information elements are associated with internal elements rather than with the entire tig. The following elements shall be used to accommodate embedding within the <ntig>: <termGrp>, <termNoteGrp>, <descripGrp>, and <adminGrp>. The attribute *lang* is required, unless inherited.
<term>	Shall contain a single-word or multi-word term, or a symbolic designation regarded as a technical term.
<termGrp>	Shall contain a <term> element and possibly at least one nested element in addition to the term.
<termNote>	Shall contain term-related information. Attributes include: *type*, which classifies the <termNote> as per the data categories specified in Annex A, A.2 and A.3.
<termNoteGrp>	Shall contain a <termNote> element and possibly at least one nested element in addition to the term-related information. Shall be used to accommodate an additional level of embedding inside of the <termGrp> element.
<descrip>	Shall contain descriptive information such as a definition, context or explanation describing concepts and terms. Attributes include: *type*, which classifies the <descrip> as per the data categories specified in Annex A, A.4 - A.7.
<descripGrp>	Shall contain a <descrip> element and possibly at least one nested element in addition to the descriptive information.
<admin>	Shall contain administrative data. Attributes include: *type*, which classifies the <admin> as per the data categories specified in Annex A, A.9-A.10.
<adminGrp>	Shall contain an <admin> element and possibly at least one nested element in addition to the administrative information.

Table 1, continued

<date>	Shall contain a single date of the format YYY-MM-DD, with the option for date-time notation as YYYY-MM-DD hh:mm:ss. Attributes include: *type*, which classifies the <date> as per data categories specified in Annex A, A.10.2.1.
<note>	Shall contain a note or annotation as comment relating to either an entire <termEntry>, an entire <tig> or <ntig> or one of the <..Grp> elements.
<descripNote>	Shall be used for note-type information used in <descripGrp> when the content of the note consists of a pick list.
<adminNote>	Shall be used for note-type information used in <adminGrp> when the content of the note consists of a pick list.
<ptr>	Shall consist of a pointer to another location in the current document. Attributes include: *type*, which classifies the <ptr> as per Annex A, A.12. *target*, which specifies the destination of the reference as one or more SGML identifiers. **NOTE –** The <ptr> GI cannot be associated with supplemental text as content of the element, as it consists solely of a start-tag with an embedded target. The <ptr>, <ref>, and <xref> elements are all considered to be *links* because they connect their current location to another targeted location within a document or to a location external to the document.
<ref>	Shall define a reference to another location in the current document, in terms of one or more identifiable elements. The <ref> GI is associated with supplemental text as content of the element, hence it consists of a start-tag with an embedded target, followed by the associated text, and closed by an end-tag. Attributes include: *type*, which classifies the <ref> as per Annex A. *target*, which specifies the destination of the reference as one or more SGML identifiers.
<xref>	Shall define a reference to a graphic, illustration, figure, table, or other external document or file using an extended pointer notation as the value of the *target* attribute of <xref>, e.g., <xref target= 'documentIdentifier'>, where the id value 'documentIdentifier' is a code for the targeted document. The user shall document the extended pointer notation that is being used by including an appropriate comment in the <encodingDesc> element of the DTD header.

Table 1, continued

	Attributes include: *type*, which classifies the external reference as per Annex A. *target*, which specifies the destination of the reference as one or more SGML identifiers. **NOTE –** External elements targeted by <xref> must be accessible to the target system for import purposes.
<hi>	Shall be used to mark a word or phrase as graphically highlighted in contrast to the surrounding text. Attributes include: *type*, which classifies the highlighted material as per Annex A. *target*, which specifies the destination of the reference as one or more SGML identifiers. NOTE – In terminology management, a major use of <hi> is to set off entailed terms, i.e., terms used in a definition, note, or other textual material that are defined elsewhere in the terminology resource. See also Annex A, A.2.2..
<foreign>	Shall identify a word or phrase as belonging to some language other than that of the surrounding text. Attributes include: *lang*, which identifies the language of the word or phrase marked.
<refObjectList>	Shall be used in the back matter and shall contain one or more back-matter objects, especially shared resources such as bibliographical entries, responsibility entries, namespace identifiers (URLs and FPIs), frequently referenced textual material, geographical location lists, external files, and the like. Attributes include: *type*, which classifies the <refObjectList> as per data categories specified in Annex A, A.11.4.1.
<refObject>	Shall contain an entry generally consisting of a shared resource such as bibliographic or responsibility information, a namespace identifier, frequently referenced textual material, an item of geographical information, a reference to an external file, and the like. Bibliographic entries should reside in the back matter or in an external document (in which case the bibliographic entry shall be referenced from the back matter using the <xref> element). NOTE – Some terminology documents contain full bibliographic entries in undifferentiated format as the content of the *source* data category (see ISO 12620:1999, A.10.19). This practice encourages redundancy and increased effort for data maintenance. This information should be converted to back-matter items if possible.

Table 1, continued

	Attributes include: *type*, which classifies the <refObject> as per data categories specified in Annex A, A.11.4.2. Unless otherwise specified, the type of <refObject> is inherited from the type of its respective <refObjectList>.
<itemSet>	Shall be used in the back matter and shall contain one or more individual items that are traditionally grouped together, e.g., the items *author's surname* and *author's first name* shall be grouped together in an <itemSet> of *type=author*. Attributes include: *type*, which classifies the <itemSet>, primarily according to the data categories listed in Annex B. This International Standard does not, however, specify the full range of other data categories that can be used with <itemSet>.
<item>	Shall contain an individual instance of back-matter information. Attributes include: *type*, which classifies information primarily as per the data categories specified in Annex B for bibliographic information. This International Standard does not, however, specify the full range of other data categories that can be used with <item>.
<itemGrp>	Shall contain one or more <item>s together with a <ptr>, <ref>, or <note>. Attributes include: *type*, which classifies information primarily as per the data categories specified in Annex B for bibliographic information. This International Standard does not, however, specify the full range of other data categories that can be used with <itemGrp>.

Figures 2 and 3 provide a schematic representation of the MARTIF full form tag name and of a full MARTIF element, respectively.