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

*Applications informatiques en terminologie — Format de transfert de
données terminologiques exploitables par la machine (MARTIF) —
Transfert négocié*



Contents

Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
4 Structuring terminological information	2
5 Terminological entries in MARTIF	4
5.1 Data categories	4
5.1.1 Specification of data categories	4
5.1.2 MARTIF tags	4
5.1.3 MARTIF attributes	11
5.1.4 Values of the attribute <i>type</i>	12
5.2 MARTIF entry structures	13
5.2.1 MARTIF document structure	13
5.2.2 The terminological entry	13
5.2.3 Treatment of quasi-equivalents	15
5.2.4 Rules governing the <termEntry>	18
5.2.5 Links	22
6 Character encoding and the <i>lang</i> attribute	23
7 Interchange procedures	24
8 The Document Type Definition (DTD) for MARTIF	24
8.1 The overall structure of terminology documents	24
8.2 Prolog	29
8.2.1 Prolog declarations	29
8.2.2 MARTIF framework	31
8.2.3 MARTIF body	31
8.2.4 MARTIF character entities	32
8.3 MARTIF header	33
8.4 MARTIF text	34
8.4.1 MARTIF components	34
8.4.2 MARTIF front	34
8.4.3 MARTIF body	34
8.4.4 MARTIF back	36
8.5 Validation	41

© ISO 1999

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 the publisher.

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

Printed in Switzerland

ANNEX A (normative): Normalized data category representation	47
ANNEX B (informative): Markup of bibliographic entries	109
ANNEX C (informative) : Data categories listed according to associated Generic Identifiers (GIs) and attributes	121
ANNEX D (informative): Data modeling variance	127
ANNEX E (informative): Sample MARTIF document	129
ANNEX F (informative): Terms and definitions taken from ISO 1087-2 and from ISO 8879:1986	133
ANNEX G (informative): Contacts for further information	135
Bibliography	136
Index	137
Index of data categories and links listed in Annex A	141

Figures, tables and examples cited in this standard

Figure 1	Sample terminological entry	3
Figure 2	Example of MARTIF full form tag names	9
Figure 3	Example of a MARTIF element	10
Figure 4	The basic structure of a MARTIF terminological entry	14
Figure 5	Basic components of a MARTIF document	25
Figure 6	Structure of the document instance	25
Figure 7	SGML Declaration	29
Figure 8	ISO 646 subset	32
Figure 9	The MARTIF framework DTD fragment	41
Figure 10	The MARTIF body DTD fragment	44
Figure 11	A sample MARTIF character entity DTD fragment	46
Table 1	MARTIF tags and their description	4
Table 2	List of MARTIF attributes	11
Table A.1	Data category classification	47
Table A.2	Interpretation for Table A.3	51
Table A.3	MARTIF data category representation	53
Table B.1	Bibliographic data categories	110
Example 1	Use of the attribute <i>type</i>	12
Example 2	Use of the attribute <i>lang</i>	12
Example 3	Full MARTIF term entry	15
Example 4	Treatment of quasi-equivalents	16
Example 5	<tig> entry	19
Example 6	<ntig> entry with use of <langSet>	19
Example 7	Use of <note> and notes on notes	20
Example 8	Use of <ptr> and <ref>	22
Example 9	Use of <foreign>	23
Example 10	Sample MARTIF document	26
Example 11	Back-matter representation	38
Example 12	Namespace identifiers	39
Example 13	Responsibility entry	40
Example 14	Responsibility references	40
Example A.1	Concept system	108
Example A.2	Thesaurus entry	108
Example B.1	Traditional bibliographic notation (presentational markup)	119
Example B.2	Sample notation according to ISO 12083	119
Example B.3	Sample notation according to ISO 12200-1	120
Example E.1	MARTIF Document No. 2	129

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.

DATA CATEGORY	DATA CATEGORY CONTENT
Subject field	appearance of materials
English term	opacity
Grammatical information, part of speech, English term	noun
English definition	degree of obstruction to the transmission of visible light
English source identifier	ASTM Standard E284
English responsibility	ASTM Technical Committee E12
German term	Opazität
Grammatical information, part of speech, German term	noun
Grammatical information, gender, German term	feminine
Definition, German	Maß für die Lichtundurchlässigkeit
German source identifier	DIN 6730:1996-05
German responsibility	Normenausschuß Papier und Pappe (NPa) im DIN
French term	opacité
Grammatical information, part of speech, French term	noun
Grammatical information, gender, French term	feminine
French definition	rapport du flux lumineux incident au flux lumineux transmis ou réfléchi par un noircissement photographique
French source identifier	HJdi1986-539
French responsibility	C.I.R.A.D.

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: <i>type</i> , 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 <i>lang</i> 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 <i>lang</i> 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 <i>lang</i> 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: <i>type</i> , 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: <i>type</i> , 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: <i>type</i> , 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 YYYY-MM-DD, with the option for date-time notation as YYYY-MM-DD hh:mm:ss. Attributes include: <i>type</i>, 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: <i>type</i>, which classifies the <ptr> as per Annex A, A.12. <i>target</i>, 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 <i>links</i> 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: <i>type</i>, which classifies the <ref> as per Annex A. <i>target</i>, 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 <i>target</i> 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: <i>type</i>, which classifies the external reference as per Annex A. <i>target</i>, 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: <i>type</i>, which classifies the highlighted material as per Annex A. <i>target</i>, 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.2.2.2.
<foreign>	Shall identify a word or phrase as belonging to some language other than that of the surrounding text. Attributes include: <i>lang</i>, 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: <i>type</i>, 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 <i>source</i> 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: <i>type</i>, 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 <i>author's surname</i> and <i>author's first name</i> shall be grouped together in an <itemSet> of <i>type=author</i>. Attributes include: <i>type</i>, 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: <i>type</i>, 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: <i>type</i>, 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.

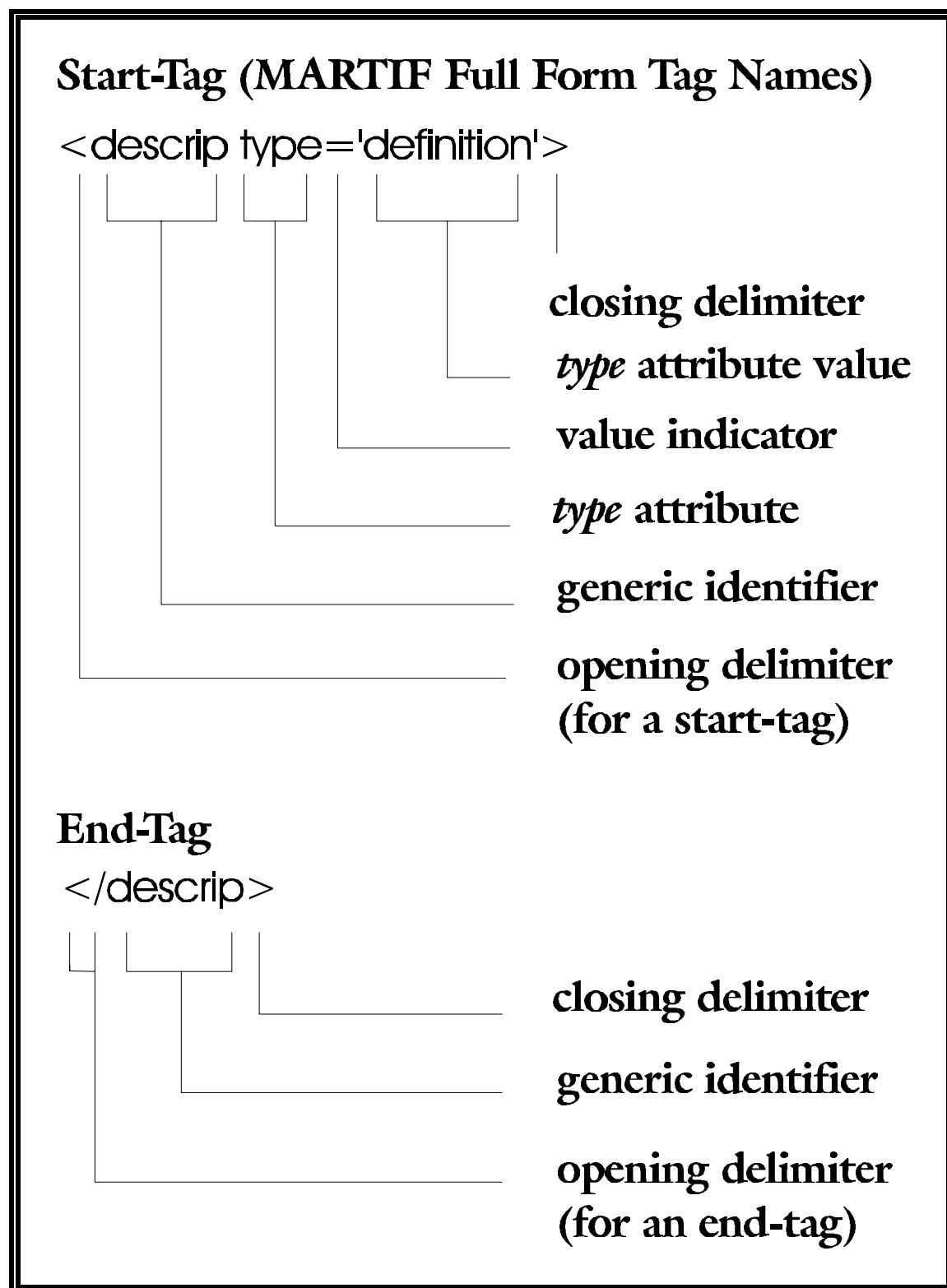


Figure 2 - Example of MARTIF full form tag names

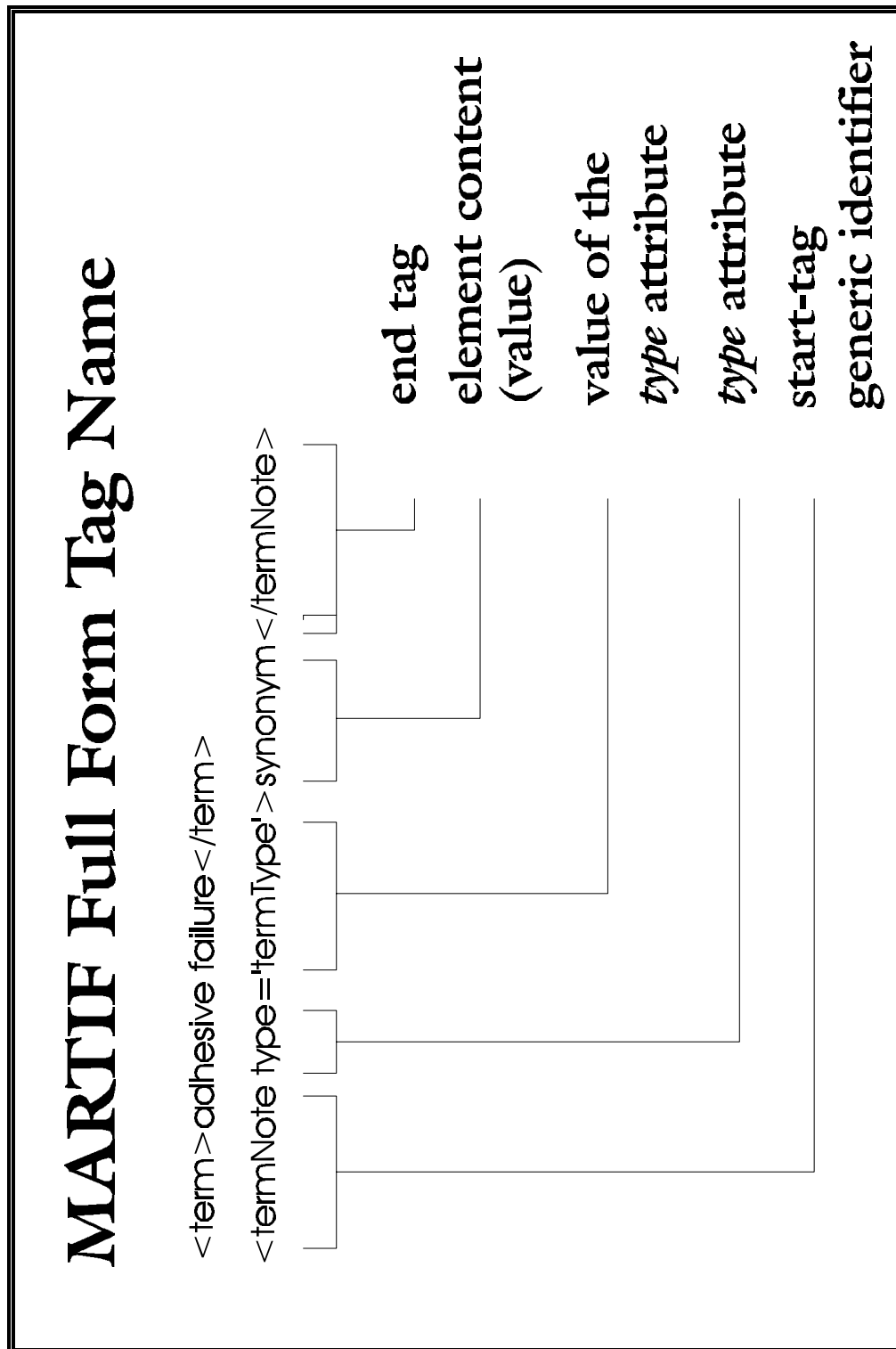


Figure 3 - Example of a MARTIF Element

5.1.3 MARTIF attributes

The attributes listed in Table 2 shall be used when it is necessary to qualify Generic Identifiers in a MARTIF document. Global attributes can be used with any MARTIF GI. Every element shall have an explicit or inherited *lang* attribute (see Clause 6). The attribute *type* shall be associated only with those GIs listed in Table 1 that specify the attribute *type*.

Table 2: List of MARTIF attributes

Global Attributes:

- id*** Shall be the unique identifier of an element.
According to ISO 8879, the ID value of an *id* shall consist of an alpha character followed by a combination of alpha characters, digits, hyphens, or dashes. Each value of *id* shall be unique within a given document.
NOTE – If necessary, any element can be assigned an *id*, although it is more frequently the case that <termEntry>s are assigned *ids*. Values of *id* used in the examples in this International Standard are arbitrarily selected. Aside from the rules stated in this attribute specification, there are no uniform criteria for assigning meaningful *id* values in terminology databases. See also Annex A, A.10.14.
- lang*** Shall indicate the language of the element content. The first two or three characters of the value of *lang* shall consist of two or three-letter symbols taken from ISO 639 or ISO 639-2 respectively (see clause 6).
Every MARTIF document shall include the *lang* attribute after the <martif> GI in the document header, e.g., <martif lang=en>. This declaration specifies the default language of the document (e.g., the language of concept-level definitions, notes, etc.). Elements contained within other elements shall automatically inherit the language of the higher element unless otherwise marked. The *lang* attribute shall be used with <langSet> or with <tig> and <ntig> unless inherited, although explicit use of *lang* is recommended for clarity in multilingual collections. It shall be used with any other element whose language differs from the language of the element in which it is embedded.

Nonglobal Attribute:

- type*** Shall be used to associate a generic identifier (GI) with an attribute value in order to form the complete tag name for a data category.
NOTE – See Annexes A and B for specific use of the attribute *type* to identify data categories.
-

5.1.4 Values of the attribute *type*

When GIs are used to represent data categories, the GIs `<term>` and `<note>` shall be used independently. As shown in Annexes A and B, other data category names shall be formed by combining four components (see Figures 2 and 3):

- a generic identifier
- the attribute *type*
- the = symbol [called a *value indicator* in SGML]
- a value enclosed in matched pairs of single or double quotation marks

No individual instance of a GI shall be used with more than one data category.

NOTE 1 – Example 1 provides the following three elements of information about the term *MARTIF*: it is an acronym (line 4), an admitted term (line 5), and a term taken from ISO 12200 (line 8). The German `<ntig>` in Example 7 in 5.2.4 illustrates a further instance of embedding involving the use of `<descripGrp>` and embedded `<ref>` qualifiers (Lines 57 and 60).

EXAMPLE 1: Use of the attribute *type*

```

1 <ntig lang=en>
2   <termGrp>
3     <term>MARTIF</term>
4     <termNote type='termType'>acronym</termNote>
5     <termNote type='normativeAuthorization'>admitted term</termNote>
6     <termNote type='timeRestriction'>Replaced E-TIF (Electronic Terminology
7       Interchange Format) as of August 1995</termNote>
8     <ptr type='sourceIdentifier' target='ISO12200.1999'>
9   </termGrp>
10 </ntig>
```

In cases where the *lang=* attribute is used or required, its content shall begin with a two or three letter language symbol taken from ISO 639 or ISO 639-2 (see Table 2 and Example 2).

EXAMPLE 2: Use of the attribute *lang*

```
<tig lang=fr> or <tig lang=fra>
```

NOTE 2 – Annex A lists the data categories taken from ISO 12620, together with their MARTIF full form tag names. ISO 12620 defines approximately two hundred data categories and permissible instances used as their content. Tags and attributes are specified in the DTD, but the values of the attribute *type* are not listed because it is desirable that the list remain open to accommodate the need for new data categories. Annex C lists data categories classified according to the generic identifiers and specific attributes with which they are associated.

5.2 MARTIF entry structures

5.2.1 MARTIF document structure

MARTIF documents shall possess a structure that conforms to ISO 8879 and to the MARTIF DTD specified in clause 8. This sub-clause discusses the structure of terminological entries (see Figure 4).

5.2.2 The terminological entry

As illustrated in Example 3, a terminological entry shall be introduced by the `<termEntry>` tag and shall contain one or more terms marked with the `<term>` tag. A single term and its associated data categories (e.g., `<termNote>`, `<descrip>`, `<admin>`), etc.) constitute a terminological information group. If all these elements function on the same level, the terminological information group is enclosed in a `<tig>` element. If additional sub-elements need to be embedded in any of the primary elements, the nested element `<ntig>` shall be used, together with the Group elements `<termGrp>`, `<descripGrp>`, or `<adminGrp>`. The element `<termNoteGrp>` is used when necessary to embed a second level of information inside a `<termGrp>`. A `<termEntry>` can be made up of a single `<tig>` or `<ntig>` or of a mixture of two or more `<tig>`s or `<ntig>`s. Multiple `<tig>`s or `<ntig>`s in any given language shall be grouped together in a `<langSet>`.

As stated above, each term shall occupy a `<tig>` or `<ntig>`, and if at all feasible, all of the `<tig>`s or `<ntig>`s associated with a concept should be contained in one `<termEntry>`. Any element that pertains specifically to information within one of the `<...Grp>` elements that does not pertain to the entire `<tig>` or `<ntig>` shall be embedded inside the respective `<...Grp>` element. Any information that pertains to the entire `<termEntry>` shall appear before the first `<tig>` or `<ntig>`. These principles are illustrated by the sample `<termEntry>` shown in Example 3.

NOTE – Example 3 represents the same data contained in Figure 1, but as a MARTIF `<termEntry>` consisting of three `<ntig>`s (English, German, and French, respectively), preceded by a *subjectField* data category that applies to the entire `<termEntry>` (Line 2). The same data are shown in Example E.1 of Annex E as a complete MARTIF document. White space (indentation, blank lines, etc.) has been used throughout the examples in this International Standard to facilitate reader understanding. Such presentational conventions are undesirable in actual data marked up for interchange, especially white space inside a character string constant.

Instead of the accented characters used in the German and French terminology information groups shown in Example 1, Example 3 uses special character strings called *character entities* (see e.g., line 14, `¨` for `ä` and line 17, `ü` for `ü`). This convention is explained in detail in 8.2.4.

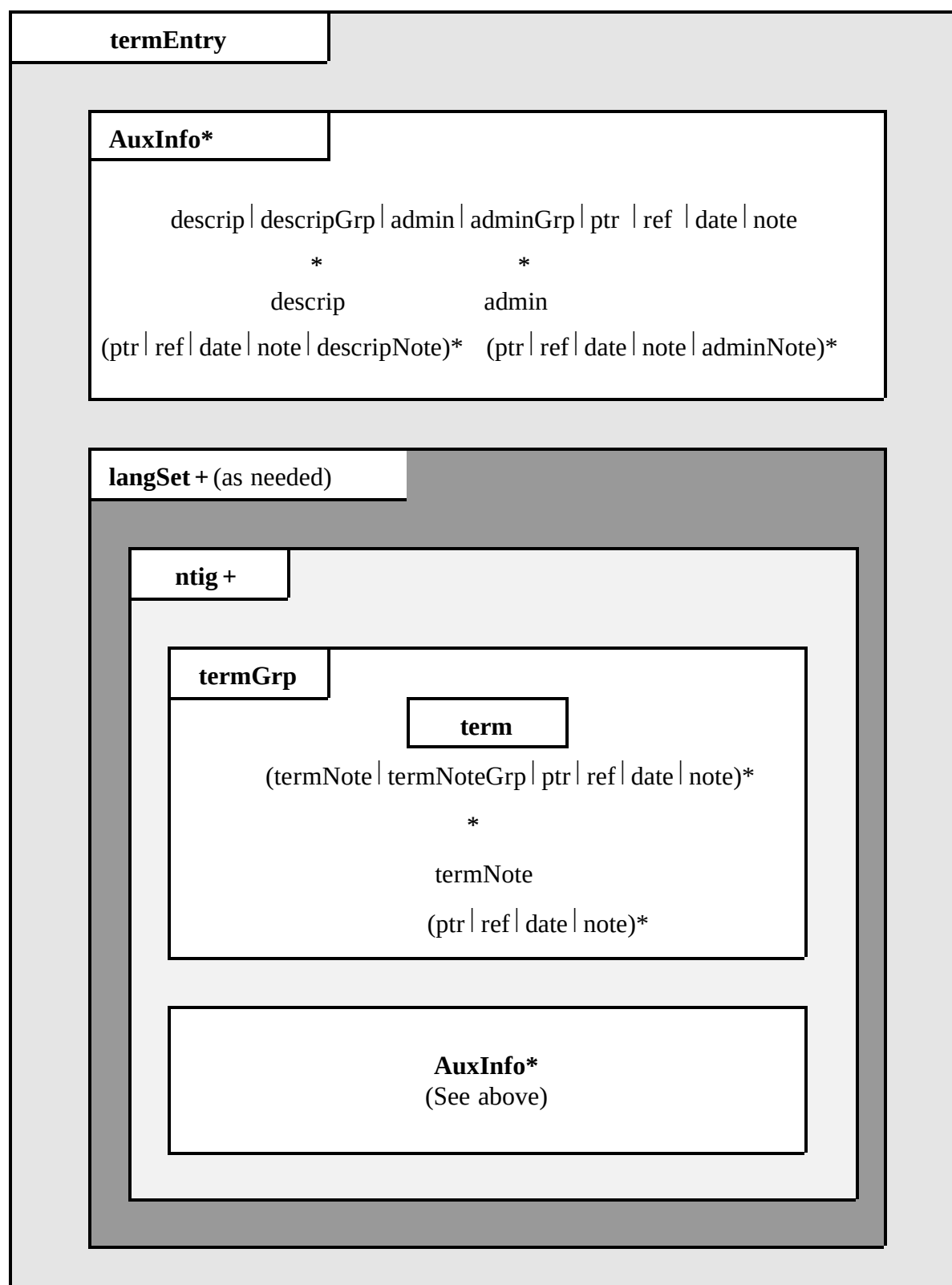


Figure 4 — The basic structure of a MARTIF terminological entry

EXAMPLE 3: Full MARTIF term entry

```

1 <termEntry id='ID000073578'>
2   <descrip type='subjectFieldLevel1'> appearance of materials </descrip>
3
4   <ntig lang=en>
5     <termGrp><term>opacity</term>
6     <termNote type='partOfSpeech'>n</termNote></termGrp>
7     <descripGrp><descrip type='definition'>degree of obstruction to the transmission of
8       visible light</descrip><ptr type='sourceIdentifier' target='ASTM.E284'></descripGrp>
9     <adminGrp><admin type='responsibility'>E12</admin> </adminGrp>
10  </ntig>
11
12  <ntig lang=de>
13    <termGrp><term> Opazit&auml;t</term>
14    <termNote type='partOfSpeech'>n</termNote>
15    <termNote type='gender'>f</termNote></termGrp>
16    <descripGrp><descrip type='definition'>Ma&szlig; f&uuml;r die
17      Lichtundurchl&auml;ssigkeit </descrip><ref type='sourceIdentifier' target='DIN-
18      6730.1996-05'>p. 383</ref></descripGrp>
19    <adminGrp><admin type='responsibility'>Normenausschu&szlig; Papier und Pappe
20      (NPa) im DIN Deutsches Institut f&uuml;r Normung e.V. </admin></adminGrp>
21  </ntig>
22
23  <ntig lang=fr>
24    <termGrp><term>opacit&eacute;</term>
25    <termNote type='partOfSpeech'>n</termNote>
26    <termNote type='gender'>f </termNote> </termGrp>
27    <descripGrp><descrip type='definition'>rapport du flux lumineux incident au flux
28      lumineux transmis ou r&eacute;fl&eacute;chi par un noircissement
29      photographique</descrip>
30    <ptr type='sourceIdentifier' target='HJdi1986'></descripGrp>
31    <adminGrp><admin type='responsibility'>C.I.R.A.D.</admin> </adminGrp>
32  </ntig>
33 </termEntry>

```

5.2.3 Treatment of quasi-equivalents

Despite the rule that all <tig>s associated with a single concept shall occupy the same <termEntry>, there are cases where it can be argued that the individual terms under consideration apply to more than one very closely related concept. In such cases, it is permissible to treat each term in its own <termEntry>, and for related entries to be joined by links (<ptr>s or <ref>s) where appropriate. The target of the link shall be the unique *id* assigned to the targeted element. For instance, Example 4 contains three <termEntry>s, each for a concept that is not quite equivalent to the other concepts. These entries are identified by the respective *id* values 'QAen01', 'Q Afr01', and 'QAde01', respectively (starting on lines 1, 24, and 51). Each entry includes a <ptr> linking it to the other entries. The <hi> element can be used to highlight entailed terms, i.e., terms that are defined in other entries, and at the same time to link to those entries.

EXAMPLE 4: Treatment of quasi-equivalents

```

1 <termEntry id='QAen01'>
2   <descrip type='subjectFieldLevel1'>quality assurance</descrip>
3   <ptr type='quasiEquivalentTerm' target='QAfr01'>
4   <ptr type='quasiEquivalentTerm' target='QAde01'>
5
6   <ntig lang=en>
7     <termGrp><term>inspection</term>
8     <termNote type='partOfSpeech'>n</termNote></termGrp>
9
10    <descripGrp><descrip type='definition'>activity such as measuring, examining testing
11    or gauging one or more characteristics of an <hi type='entailedTerm'
12    target='QAen1.1'>entity</hi> and comparing the results with specified requirements in
13    order to establish whether <hi type='entailedTerm' target='QAen2.9'>conformity</hi> is
14    achieved for each characteristic</descrip>
15    <ptr type='sourceIdentifier' target='ISO8402.en2.15'>
16    </descripGrp>
17
18    <adminGrp><admin type='responsibility'>ISO TC 176</admin>
19    </adminGrp>
20  </ntig>
21
22 </termEntry>
23
24 <termEntry id='QAfr01'>
25   <descrip type='subjectFieldLevel1'>quality assurance</descrip>
26   <ptr type='quasiEquivalentTerm' target='QAen01'>
27   <ptr type='quasiEquivalentTerm' target='QAde01'>
28
29   <ntig lang=fr>
30     <termGrp><term>contrôle</term>
31     <termNote type='partOfSpeech'>n</termNote>
32     <termNote type='gender'>f</termNote></termGrp>
33
34     <descripGrp><descrip type='definition'>activités telles que mesurer, examiner,
35     essayer ou passer au calibre une ou plusieurs caractéristiques d'une <hi
36     type='entailedTerm' target='QAfr1.1'>entité</hi> et comparer les
37     résultats aux exigences spécifiées en vue de déterminer si
38     la <hi type='entailedTerm' target='QAfr2.9'>conformité</hi> est obtenue pour
39     chacune de ces caractéristiques</descrip>
40     <ptr type='sourceIdentifier' target='ISO8402-fr2.15'>
41     <note>En français, le terme 'inspection' peut désigner une
42     activité de surveillance de la qualité; conduite dans le cadre d'une
43     mission bien définie.</note>
44     </descripGrp>
45

```

46 <adminGrp><admin type='responsibility'>ISO TC 176</admin> </adminGrp>
47 </ntig>
48
49 </termEntry>
50
51 <termEntry id='QAde01'>
52 <descrip type='subjectFieldLevel1'>quality assurance</descrip>
53 <ptr type='quasiEquivalentTerm' target='QAfr01'>
54 <ptr type='quasiEquivalentTerm' target='QAen01'>
55
56
57 <ntig lang=de>
58 <termGrp><term>Prüfen</term>
59 <termNote type='partOfSpeech'>n</termNote>
60 <termNote type='gender'>n</termNote></termGrp>
61
62 <descripGrp><descrip type='definition'>feststellen, ob der Prüfgegenstand
63 (Probekörper, Probe, Meßgerät) eine oder mehrere vereinbarte oder
64 vorgeschriebene oder erweiterte Bedingungen erfüllt, insbesondere ob
65 vorgegebene Fehlergrenzen oder Toleranzen eingehalten werden</descrip>
66 <note>Mit dem Prüfen ist daher immer eine Entscheidung verbunden. Das
67 Prüfen kann subjektiv durch Sinneswahrnehmung ohne Hilfsgerät oder
68 objektiv mit Meßgeräten oder mit Prüfgeräten, die auch
69 automatisch arbeiten können, geschehen. Ein subjektives Prüfen
70 führt meist nur zu einer qualitativen Angabe.</note>
71 <ref type='sourceIdentifier' target='FHtb1985'>p. 405-406</ref>
72 <descrip type='transferComment'>German 'Prüfen' encompasses both English
73 'inspection' and 'test'.</note>
74 </descripGrp>
75 <adminGrp><admin type='responsibility'> Normenausschuß
76 Qualitätsmanagement Systeme und angewandte Statistik im DIN Deutsches
77 Institut für Normung e.V.</admin></adminGrp>
78 </ntig>
79
80 </termEntry>

5.2.4 Rules governing the <termEntry>

The following rules apply to a <termEntry>

1. One <termEntry> shall be created for one (i.e., for each individual) concept.
2. Each term, synonym, variant, etc. shall occupy its own <tig> or <ntig>, with appropriate cross-references if necessary. (See Example 5 for the use of <tig> and Example 6 for the use of <ntig>.)
3. The normalized mode of the data categories shall be used (see normative Annex A).
4. If any element refers to the entire <termEntry> and not just to one <langSet>, <tig> or <ntig>, it shall be placed after the <termEntry> tag and before the start tag for the first <langSet>, <tig> or <ntig>.
5. If, for instance, a <note>, <termNote>, or link exists that refers to an individual element (such as to a <term>, etc.), but not to the entire terminological information group, the <ntig> element shall be used, together with the appropriate Group element (<termGrp>, <descripGrp> or <adminGrp>). The item in question shall be enclosed in the Group element. Both <tig> and <ntig> elements can be used together in the same <termEntry>.
6. In the event that an additional note or link shall be referenced to one of the data categories introduced by the <termNote> tag, <termNoteGrp> shall be introduced into the <termGrp> element, together with <termNote> and the respective <note> (see Example 6).
7. If an element must be referenced to another element embedded inside one of the <... Grp> elements, but the first element does not pertain to the entire group, the reference shall be made using the <ref> tag and the referenced information shall be included as the content of the <ref> element (see Example 7).
8. Standard values for language symbols (*lang=*) shall be used as specified in ISO 639 and ISO 639-2. Further specification of regional variation and writing system declarations (WSD) shall be indicated as set forth in Clause 6.
9. The *lang* attribute shall appear with the <martif> element in the header of every MARTIF document in order to specify the default language of the document. Multiple <ntig>s in a given language shall be enclosed in a <langSet>, which shall contain a *lang=* attribute identifying the language of the <langSet> if this language differs from the default language. Freestanding <tig>s or <ntig>s shall also contain a *lang=* attribute if their language differs from the default language. Sub-elements within <tig> and <ntig> inherit their respective language unless the sub-element is accompanied by an explicit *lang* attribute or is contained in another element with its own *lang* identifier.
10. Standardized values for dates shall be used as specified in ISO 8601 (e.g., 1995-10-30, with the possibility of expansion to date and time, e.g., 1995-10-30 12:32:41 where needed.). See Annex A, A.10.2 for data category names of administrative dates used in terminology management.

Example 5 illustrates a simple <tig> without the inclusion of embedded information (see Rule 2).

EXAMPLE 5: <tig> entry

```

1 <termEntry id='ASTMD1517-053'>
2
3   <tig lang=en>
4     <term>fancy leather</term>
5     <descrip type='definition'>leather made from hides and skins of all kinds that have
6       commercial importance and value primarily because of grain, or distinctive finish,
7       whether natural or the result of processing</descrip>
8     <note>Such processing may be graining, printing, embossing, ornamenting (including
9       gold, silver, and aluminum finishes), or any other finishing operation used to enhance
10      the appeal of the leather.</note>
11     <admin type='responsibility'>ASTM D1517</admin>
12   </tig>
13
14 </termEntry>

```

Example 6 illustrates the way that <langSet> can be used to enclose multiple <ntig>s for terms in the same language (lines 15-36). In the <ntig> for the French word *inspection* shown here, there is a note referenced to a <termNote type='geographicalUsage'>, which necessitates the use of a <termNoteGrp> inside <termGrp> (line 29, ff). As this example focuses on the use of <termNote> and <termNoteGrp>, other elements have been represented by ellipses in the term entry (see Rules 2 and 5).

EXAMPLE 6: <ntig> entry with use of <langSet>

```

1 <termEntry id='QA2.556'>
2   <descrip type='subjectFieldLevel1'>quality assurance</descrip>
3   <ref type='sourceIdentifier' target='jbQA1994'>p. 345</ref>
4
5   <langSet lang=en>
6     <ntig>
7       <termGrp>
8         <term>inspection</term>
9         <termNote type='partOfSpeech'>n</termNote>
10      </termGrp>
11      ...
12    </ntig>
13  </langSet>
14
15  <langSet lang=fr>
16    <ntig>
17      <termGrp>
18        <term>contr&ocirc;le</term>
19        <termNote type='gender'>f</termNote>
20        <termNote type='geographicalUsage'>FR</termNote>
21      </termGrp>
22      ...
23    </ntig>

```

```

24
25 <ntig>
26   <termGrp>
27     <term>inspection</term>
28     <termNote type='gender'>f</termNote>
29     <termNoteGrp><termNote type='geographicalUsage'>CA</termNote>
30     <note>Although an earlier standard cited the Canadian usage, the current standard
31       has eliminated this reference.</note>
32   </termNoteGrp>
33 </termGrp>
34   ...
35 </ntig>
36 </langSet>
37
38 </termEntry>

```

In Example 7, the notes that follow the <descripGrp> in each parallel <ntig> pertain to the term treated in the <ntig> (Lines 10, 26, and 48 ff.). The information included in the <ref> element that references the note in the German <ntig> (line 56 ff.) is a note on that note, and the information in the second <ref> is a note on the note on the note (see Rule 6). Such complex data structures should be avoided unless absolutely necessary to meet user needs.

EXAMPLE 7: Use of <note> and notes on notes

```

1 <termEntry id='ISO9000-1-A1'>
2   <descrip type='subjectFieldLevel1'>quality assurance</descrip>
3   <ref type='sourceIdentifier' target='ONORM-ISO9000-1'>p. 34</ref>
4
5   <ntig lang=en>
6     <termGrp><term>quality</term>
7     <termNote type='partOfSpeech'>noun</termNote></termGrp>
8     <descripGrp><descrip type='definition'>totality of characteristics of an entity that bear
9       on its ability to satisfy stated and implied needs</descrip>
10    <note id='ISO9000-1A1en21'>The term 'quality' is not used as a single term to express
11      a degree of excellence in a comparative sense, nor should it be used in a quantitative
12      sense for technical evaluations. To express these meanings, a qualifying adjective
13      should be used. For example, use can be made of the following terms:
14      a) 'relative quality' where ...
15      b) 'quality level' where ...</note></descripGrp>
16    <adminGrp><admin type='responsibility'>ISO TC 176</admin>
17    </adminGrp>
18  </ntig>
19
20  <ntig lang=fr>
21    <termGrp><term>qualit&eacute;</term>
22    <termNote type='gender'>m</termNote></termGrp>
23    <descripGrp><descrip type='definition'>ensemble des caract&eacute;ristiques d'une
24      entit&eacute; qui lui conf&egrave;rent l'aptitude &agrave; satisfaire des besoins
25      exprim&eacute;s et implicites</descrip>

```

26 <note id='ISO9000-1A1fr21'>Il convient que le terme «qualité»
 27 ne soit utilisé isolément ni pour exprimer un degré d'excellence
 28 dans un sens comparatif, ni pour des évaluations techniques dans un sens
 29 quantitatif. Pour exprimer ces deux sens, il est bon qu'un adjectif qualificatif soit
 30 utilisé. Par exemple, on peut employer les termes suivants:
 31 a) «qualité relative» lorsque . . .
 32 b) «niveau de qualité» dans un sens quantitatif ...</note>
 33 </descripGrp>
 34 <adminGrp><admin type='responsibility'>ISO TC 176
 35 </admin></adminGrp>
 36 </ntig>
 37 <ntig lang=de>
 38 <termGrp><term>Qualität</term>
 39 <termNote type='gender'>f</termNote></termGrp>
 40 <descripGrp>
 41 <descrip type='definition'>Gesamtheit von Merkmalen (und Merkmalswerten) einer
 42 Einheit bezüglich ihrer Eignung, festgelegte und vorausgesetzte Erfordernisse
 43 zu erfüllen</descrip>
 44 <note>Fußnote in der deutschsprachigen Fassung: 'Festgelegte und
 45 vorausgesetzte Erfordernisse' sind zwei spezifische Konkretisierungen ...
 46 </note>
 47 <note id='ISO9000-1A1de21'>Die Benennung 'Qualität' sollte weder als
 48 einzelnes Wort gebraucht werden, um einen Vortrefflichkeitsgrad im vergleichenden
 49 Sinn auszudrücken, noch sollte sie in einem quantitativen Sinn f&r technische
 50 Bewertungen verwendet werden. Um diese Bedeutung auszudrücken, sollte
 51 ein qualifizierendes Adjektiv benutzt werden. z.B. können folgende
 52 Benennungen verwendet werden:
 53 a) 'Relative Qualität', wo . . .
 54 b) 'Qualitätslage' in einem quantitativen Sinne . . .</note>
 55 <ref type='note' id='ISO9000-1A1en21sub1' target='ISO9000-
 56 1A1de21'>Fußnote in der deutschsprachigen Fassung: An diesen Stellen
 57 weicht die Originalfassung . . . ab.</ref>
 58 <ref type='sourceIdentifier' target='ISO9000-1A1en21sub1'>p. 33</ref>
 59 </descripGrp>
 60 <adminGrp><admin type='responsibility'>DIN/ONORM/ SNV/ISO TC 176
 61 committees</admin></adminGrp>
 62 </ntig>
 </termEntry>

5.2.5 Links

Terminology documents can utilize a variety of cross-references between <termEntry>s, for instance as illustrated by the <ptr> elements used in Example 4 in 5.2.3. Links shall be implemented using <ptr> and <ref> linking elements, together with a value of the attribute *type* to indicate the category of link that is being used.

NOTE – Links constitute relational information rather than data category information. Therefore, types of links are defined in Annex A, A.12 of this International Standard, but they are not the subject of ISO 12620. Annex C indicates data categories that can be used with <ptr> and <ref>, as well as with <hi>, to form links.

The difference between <ptr> and <ref> can be illustrated quite clearly by examining their use for linking <termEntry>s to bibliographic entries. If, as is the case with the reference to ASTM E284 in Example 8, the total source identifier is contained as the content of the *target* attribute of the link, <ptr> shall be used. If, on the other hand, a page number is included, this page number shall appear as the content of a linking element introduced by the <ref> tag.

EXAMPLE 8: Use of <ptr> and <ref>

```
<ptr type='sourceIdentifier' target='ASTM E284'>  
<ref type='sourceIdentifier' target='FHdn1983'> p. 383 </ref>
```

NOTE – These <ptr> and <ref> links point to material in the back matter. The format for the treatment of the bibliographic entries targeted by these links is discussed in 8.4.4 in the context of the MARTIF back element. They can, for instance, point to bibliographic entries contained in a <refObjectList type='bibl'>containing either bibliographic entries in the form of <refObject>s or they can point to <xref> elements in the <back> element of the MARTIF document, which in turn point to external bibliographic information in a separate SGML document encoded according to the DTD described in ISO 12083. This document must reside in the same system as the MARTIF document. The sample MARTIF document shown in Example 10 shows the second option. The sample document shown in Annex E includes complete bibliographic information in the <back> element.

MARTIF does not provide the capability to differentiate the individual elements in the bibliographic reference if the full bibliographic citation is included in the <termEntry>. In such cases, the citation shall be identified with the tag name <admin type='source'>and shall occur as a self-contained data category. Although some terminological databases do include full bibliographic information in each terminological entry, this practice can lead to redundancy and increased data management costs.

6 Character encoding and the *lang* attribute

The language of every element in a MARTIF document shall be clearly indicated. If the language of the element is the same as that of the element in which it is embedded (e.g., the default <martif> language or the language of the element's respective <langSet>, <tig>, or <ntig>), no additional explicit markup is required. This feature is commonly referred to as the *principle of inheritance*. If, however, the language of the element differs from that of its surrounding context, an explicit *lang* attribute shall be used to override the prevailing language identifier. A typical example might be the use of the <foreign> element to set off a foreign word or phrase (e.g., lines 4 and 5 in Example 9).

EXAMPLE 9: Use of <foreign>

```
1 <ntig lang=en>
2   <termGrp><term>inspection</term></termGrp>
3   <descripGrp><descrip type='definition'>...</descrip>
4   <note>The French usage of <foreign lang=fr>contrôle</foreign> corresponds to the
5     North American variant <foreign lang=fr>inspection </foreign>, which illustrates the
6     influence of English on North American French.</note></descripGrp>
7 </ntig>
```

NOTE – *lang* does not apply to the data *inside* a tag, rather only to the data *between* the tags (i.e., *after* the '>' of the start-tag to just *before* the '<' of the end-tag).

The value of the *lang* attribute shall be or begin with a two or three-letter, lowercase language code element as specified in ISO 639 or ISO 639-2. It can feature an extension consisting of uppercase country code elements as specified in ISO 3166-1 or writing system information as noted below, or both.

Character entities shall be encoded according to ISO 8879:1986, Annex D, where possible (see 8.2.4), and exceptions shall be recorded in the <encodingDesc> element of the document header (see 8.3). Each language symbol shall be associated with a writing system, which includes a set of characters as well as conventions affecting directionality, correspondence between upper and lower case, and sorting order. The <encodingDesc> element shall also be used to record all relevant information about the writing systems used in the document. In the event that more than one writing system or representational system is used to represent a language (e.g., representation in differing scripts, such as Latin and Cyrillic, or representation via transliteration, transcription, or romanization, or variations on any of these systems), an extension shall be appended to the language symbol using index numbers, e.g., ru1, ru2, ru3, etc. All extensions shall be explained in the <encodingDesc> and referenced to a standard character repertoire (e.g., ISO 10646), a set or sets of SGML character entities, or a standard system for alternate graphical representation (transliteration, transcription, or romanization system). 8.3 also provides additional information on documenting Writing System Declarations.

7 Interchange procedures

In order to carry out terminology interchange with MARTIF, terminological data shall be exported from the source database to MARTIF, usually by means of an export routine designed for this purpose. Import routines are then necessary to import MARTIF documents into target databases. When setting up any individual interchange relationship, specific features, such as system architecture and entry structure of the source and target databases, should generally be examined to determine if it will be necessary to negotiate conversion routines in order to facilitate problem-free interchange.

The data categories used shall be identified as indicated in Annex A. In assigning data category names from Annex A to data categories in the source database, users should consult ISO 12620 to ensure that data category content is harmonized with the data category definitions provided in ISO 12620.

8 The Document Type Definition (DTD) for MARTIF

8.1 The overall structure of terminology documents

The full MARTIF Document Type Definition (DTD) shall consist of the three-component DTD files represented here in Figure 9 (the framework), Figure 10 (the body), and Figure 11 (the character entities). These components shall be combined as specified in the prolog.

The overall structure of a MARTIF document shall conform to the principles laid down in ISO 8879. A complete MARTIF document shall consist of a prolog, followed by a document instance of type MARTIF (for Machine-Readable Terminology Interchange Format). The document instance shall consist of a header (`<martifHeader>`) followed by the text, which in turn consists of optional front matter, the body (a sequence of terminological entries), and optional back matter.

NOTE – Figure 5 summarizes the components of a MARTIF file (in outline form) represented as a complete SGML document. Figure 6 shows the structure of the document instance as expressed by standard SGML generic identifiers. Example 10 shows a sample MARTIF document that contains a prolog, a header, no front, a body (consisting in this case of three terminological entries), and a back.

Example 10 contains a complete MARTIF document consisting of a prolog (line 1), a `<martifHeader>` (line 7, ff.), a body (line 21, ff.) made up of three `<termEntry>`s (starting on lines 23, 58, and 89, resp.), and a back (line 124, ff.), which in this case is comprised of a `<refObjectList>` containing a single `<refObject>` consisting of an external reference to a bibliographic file. The file is cited as residing in a subdirectory for ISO 12083-conformant files, and this subdirectory shall be present on the same system as the MARTIF document in order to facilitate the external reference. The three `<termEntry>`s provide an illustration of linkage between related concepts.

- I. Prolog
- II. Document instance (<martif lang=en>)
 - A. header (<martifHeader>)
 - B. text
 - 1. front (optional)
 - 2. body
 - a. first terminological entry <termEntry>
(minimum of one)
 - b. second terminological entry <termEntry>
 - c. etc. (additional terminological entries)
 - 3. back (optional)

Figure 5—Basic components of a MARTIF document

```
<martif lang=en>
<martifHeader>
... (The header goes here.)
</martifHeader>
<text>
  <body>
    ... (The terminological entries go here.)
  </body>
  <back>
    ... (Included bibliographic entries go here.)
    ... (Any external references (<xref>s) also go here.)
  </back>
</text>
</martif>
```

Figure 6—Structure of the document instance

EXAMPLE 10: Sample MARTIF document

```

1 <!DOCTYPE martif PUBLIC "ISO 12200:1999//DTD for MARTIF (framework) //EN" [
2 <!ENTITY % mtf-body PUBLIC "ISO 12200:1999//DTD for MARTIF (body) //EN">
3 <!ENTITY % mtf-ents PUBLIC "ISO 12200:1999//ENTITIES for MARTIF (sets) //EN">
4 ]>
5
6 <martif lang=en>
7 <martifHeader>
8
9 <fileDesc>
10 <titleStmt><title>Example 10: a complete MARTIF document</title>
11 </titleStmt>
12 <publicationStmt><p>not published separately</p></publicationStmt>
13 <sourceDesc><p>from ISO DIS 12200, body: Example 10</p></sourceDesc>
14 </fileDesc>
15 <revisionDesc>
16 <change><p>1995-11-15: edited to accommodate declaration of public
17 document</p></change>
18 </revisionDesc>
19 </martifHeader>
20
21 <text>
22 <body>
23
24 <termEntry id='ISO-3534-2.32'>
25   <descrip type='subjectFieldLevel1'>statistics</descrip>
26   <ref type='sourceIdentifier' target='ISO-3534'>p. 15</ref>
27   <admin type='responsibility'>ISO TC 69</admin>
28
29   <ntig lang=en>
30     <termGrp>
31       <term>mean deviation</term>
32       <termNote type='partOfSpeech'>noun</termNote>
33     </termGrp>
34     <descripGrp>
35       <descrip type='definition'>the arithmetic mean of the deviations from an origin
36       when all deviations are given a positive sign</descrip>
37       <note>Generally, the chosen origin is the arithmetic mean, although the mean
38       deviation is minimal when the median is taken as an origin.</note>
39     </descripGrp>
40   </ntig>
41
42   <ntig lang=fr>
43     <termGrp>
44       <term>&eacute;cart moyen</term>
45       <termNote type='partOfSpeech'>noun</termNote>
46       <termNote type='gender'>m</termNote>
47     </termGrp>
48   </ntig>

```

49 <descrip type='definition'>moyen arithmétique des écarts par rapport
50 à une origine; les écarts sont pris en valeur absolue.</descrip>
51 <note>Généralement, l'origine choisie est la moyenne
52 arithmétique bien que l'écart moyen soit minimal quand on prend la
53 médiane pour origine.</note>
54 </descripGrp>
55 </ntig>
56 </termEntry>
57
58 <termEntry id='ISO-3534-2.53'>
59 <descrip type='subjectFieldLevel1'>statistics</descrip>
60 <ref type='sourceIdentifier' target='ISO-3534'>p. 19</ref>
61 <admin type='responsibility'>ISO TC 69</admin>
62
63 <ntig lang=en>
64 <termGrp>
65 <term>sampling error</term>
66 <termNote type='partOfSpeech'>noun</termNote>
67 </termGrp>
68 <descripGrp>
69 <descrip type='definition'>part of the total estimation error of a parameter due to
70 the random nature of the sample</descrip><ptr type='relatedTerm' target='ISO-
71 3534-2.56'>
72 </descripGrp>
73 </ntig>
74
75 <ntig lang=fr>
76 <termGrp>
77 <term>erreur d'échantillonnage</term>
78 <termNote type='partOfSpeech'>noun</termNote>
79 <termNote type='gender'>m</termNote>
80 </termGrp>
81 <descripGrp>
82 <descrip type='definition'>partie de l'erreur totale d'estimation d'un
83 paramètre et la valeur vraie de ce paramètre</descrip><ptr
84 type='relatedTerm' target='ISO-3534-2.56'>
85 </descripGrp>
86 </ntig>
87 </termEntry>
88
89 <termEntry id='ISO-3534-2.56'>
90 <descrip type='subjectFieldLevel1'>statistics</descrip>
91 <ref type='sourceIdentifier' target='ISO-3534'>p. 19</ref>
92 <admin type='responsibility'>ISO TC 69</admin>
93
94 <ntig lang=en>
95 <termGrp>
96 <term>standard error</term>
97 <termNote type='partOfSpeech'>noun</termNote>
98 </termGrp>
99

```

100      <descripGrp>
101          <descrip type='definition'>the standard deviation of an estimator; the standard error
102          provides an estimation of the random part of the total estimation error involved in
103          estimating a population parameter from a sample</descrip>
104      </descripGrp>
105  </ntig>
106
107  <ntig lang=fr>
108      <termGrp>
109          <term>erreur-type</term>
110          <termNote type='partOfSpeech'>noun</termNote>
111          <termNote type='gender'>m</termNote>
112      </termGrp>
113      <descripGrp>
114          <descrip type='definition'>&Eacute;cart-type d'un estimateur; l'erreur-type permet
115          une estimation de la partie aléatoire de l'erreur totale d'estimation faite en
116          estimant un paramètre de la population &agrave; partir d'un
117          échantillon</descrip>
118          <ptr type='relatedTerm' target='ISO-3534-2.53'>
119      </descripGrp>
120  </ntig>
121 </termEntry>
122 </body>
123
124 <back>
125     <refObjectList type='bibl'>
126         <refObject>
127             <item id='ISO-3534'>
128                 <xref target='c:\bibl\ISO-12083\ISO-3534BIB.NIS'>ISO 12083-conformant
129                 bibliography file</xref>
130             </item>
131         </refObject>
132     </refObjectList>
133
134 </back>
135
136 </text>
137
138 </martif>

```

8.2 Prolog

8.2.1 Prolog declarations

The prolog component in Example 10 (see 8.1) is:

```
<!DOCTYPE martif PUBLIC "ISO 12200:1999//DTD for MARTIF (framework) //EN" [
<!ENTITY % mtf-body PUBLIC "ISO 12200:1999//DTD for MARTIF (body) //EN">
<!ENTITY % mtf-ents PUBLIC "ISO 12200:1999//ENTITIES for MARTIF (sets) //EN">
]>
```

These lines, which should be essentially the same for all MARTIF documents, refer to three external files, each identified by a public name. Each of these three files is a DTD fragment, but the prolog combines them into the full MARTIF DTD.

The prolog (lines 1-4) declares that there can be any number of document instances of type <martif>. The *lang=en* attribute indicates that the default metalanguage of the document is English. The overall structure of the document instance outlined in Figure 5 is illustrated in Example 10 (see 8.1).

Depending on the default SGML declaration used by the selected parser, it can be necessary to include an SGML declaration at the beginning of the document (preceding the document type declaration statement in the prolog) in order to parse a MARTIF document. Typically, a default SGML declaration will suffice except that in the QUANTITY section of the declaration, NAMELEN shall be set to 32 or higher (whereas the default value is often only 8). This is essential because several of the generic identifiers in the MARTIF DTD are longer than 8 characters (see Figure 7).

Figure 7—SGML Declaration

```
<!SGML "ISO 8879:1986"
```

```
--
```

```
  MARTIF SGML declaration for local processing with SGMLS:
  extended capacity points (Namelen 32, Litlen 512)
  full ASCII character set instead of ISO/IEC 646 subset
```

```
--
```

```
CHARSET
```

```
BASESET "ISO 646:1983//CHARSET
```

```
  International Reference Version (IRV)//ESC 2/5 4/0"
```

```
DESCSET      0      9      UNUSED
              9      1      9
              10     1      10
              11     2      UNUSED
              13     1      13
              14    12      UNUSED
              26     1      UNUSED -- eof
```

	27	5	UNUSED
	32	95	32
	127	1	UNUSED

BASESET "ISO Registration Number 100//CHARSET
ECMA-94 Right Part of Latin Alphabet Nr. 1//ESC 2/13 4/1"

DESCSET	128	32	UNUSED
	160	95	32
	255	1	UNUSED

CAPACITY PUBLIC "ISO 8879:1986//CAPACITY Reference//EN"
SCOPE DOCUMENT

SYNTAX

SHUNCHAR CONTROLS 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
127 255

BASESET "ISO 646:1983//CHARSET
International Reference Version (IRV)//ESC 2/5 4/0"

DESCSET	0	128	0
FUNCTION	RE	13	
	RS	10	
	SPACE	32	
TAB	SEPCHAR	9	
-- EOF	SEPCHAR	26	--

NAMING

	LCNMSTRT	""
	UCNMSTRT	""
	LCNMCHAR	"-."
	UCNMCHAR	"-."
	NAMECASE	GENERAL YES
	ENTITY	NO

DELIM GENERAL SGMLREF
SHORTREF SGMLREF
NAMES SGMLREF
QUANTITY SGMLREF

ENTLVL	20
NAMELEN	32
LITLEN	512
TAGLVL	24

FEATURES

MINIMIZE	
DATATAG	NO
OMITTAG	YES
RANK	NO
SHORTTAG	YES

LINK

SIMPLE	NO
IMPLICIT	NO
EXPLICIT	NO

OTHER	
CONCUR	NO
SUBDOC	YES 10
FORMAL	YES

APPINFO
NONE
>

8.2.2 MARTIF framework

The MARTIF framework is found in Figure 9. The framework DTD declares that the overall structure of a MARTIF document shall consist of a header followed by a text (see the *martif* element declaration in Section 2 of Figure 9). The header shall provide general information about the document. The text shall consist of an optional front element, a required body element, and an optional back element (see text element declaration in Section 3 of Figure 9).

The framework DTD also declares that several low-level elements shall be used in the header, front, and body. The body shall consist of a sequence of terminological entries, such as the examples shown in this International Standard; however, these entries cannot form an SGML document by themselves.

NOTE – The framework component defines the macrostructure into which the entries fit, in analogy to the framework of a house that has only two inside supporting walls. The supporting walls divide the house into three major sections: a small "front", a large "body", and a small "back". Within this framework, the house can be remodelled to have various combinations of rooms in the body section. The low-level elements can be thought of as furniture and other fixtures. In fact, the framework component of the MARTIF DTD could be used for various types of documents. The MARTIF framework refers to two entities, mtf-body and mtf-ents. These entities are not defined in the framework DTD, but they are defined in two respective files that combine with the framework to form the complete DTD. It is critical that all these files be present on the system.

There are no system-specific identifiers in the DOCTYPE statement of the MARTIF DTD. The PUBLIC names are unique and their contents shall remain untouched, with the exception of the third component of the DTD, which can contain user-specified character entities.

8.2.3 MARTIF body

The mtf-body entity is referred to at the end of Section 3 in Figure 9 and its content is presented in Figure 10. This file defines the internal structure of the body element by declaring that the mtf-body entity shall provide the definition of the body component of a MARTIF document and that the body shall be composed of terminological data. This definition also provides the portion of the DTD that determines the structure of terminological entries. This file should not be modified because otherwise interchange can be seriously hindered.

8.2.4 MARTIF character entities

Unless otherwise indicated in the <encodingDesc> element of the header (which shall be cited as *Option 1* in this document), MARTIF documents shall use only those characters defined in ISO 646 and characters defined as character entities composed entirely of 646 characters (*Option 2*). (For instance, the word *Mqß* in the German <tig> shown in Example 3 contains the character *ß* (German *sharp s*), which is not contained in the ISO 646 subset. Consequently, this character is represented by the entity “ß”, which describes this character as “s z ligature”. Unless indicated for Option 1, all non-646 characters shall be represented with comparable character entities. Once the document has been safely transmitted to a target platform, the character entities can be automatically converted to a more convenient target-platform local representation.

In some cases, even the characters in ISO 646 are prone to corruption during transmission. In instances where interchange takes place over links not reliable for the full character set, such as character sets used in mainframe computer environments, the characters subject to misinterpretation and corruption shall be replaced by standard entities. The characters least susceptible to loss or misinterpretation in transit among systems are shown in Figure 8.

Figure 8—ISO 646 subset

```

a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9
" % & ' ( ) * + , - . / : ; < = > ? _

```

NOTE – This set also includes the space character.

A MARTIF document can be represented using only the characters in Figure 8 plus character entities composed solely of characters from Figure 8. Of the characters shown in Figure 8, the only ISO/IEC 646 characters that are absolutely essential during transmission are the letters a-z (upper and lower case), the digits 0-9, and the punctuation symbols '&' (ampersand) and ';' (semi-colon). All other characters can then be defined using character entities. When ISO 646 characters from Figure 8 are temporarily represented as entities during transmission, however, they should be reconverted to ISO/IEC 646 characters before processing a MARTIF document using an SGML parser or other SGML-aware software.

Representing all non-ISO-646 characters as character entities allows a MARTIF document to be transmitted across any transmission path to any platform that supports ISO 646. When the MARTIF document has been received and parsed, the non-ISO-646 character entities are typically converted to a local representation for convenient viewing, editing, and processing.

The framework component of the DTD allows the user to specify the list of character entities that shall be used to represent characters in a MARTIF document instance. This specification is accomplished by allowing the user to modify the content of the text entity *mtf-ents* referenced in the framework (see Figure 9, Section 1). The text entity *mtf-ents* is a metafile

that references one or more sets of character entities, each set consisting of a text entity. The sample *mtf-ents* metafile shown in Figure 11 references a public text entity called *latin1.ents* and a user-defined text entity called *extra.ents*. The *latin1* character entities are insufficient for terminology work in non-latin-alphabet languages such as Russian and Greek and non-alphabetic languages such as Chinese, Japanese, and Korean (CJK). Users should update the content of *extra.ents* as needed in order to accommodate the languages used in the terminological entries of a particular MARTIF document instance.

Unless otherwise indicated in the `<encodingDesc>` as noted above for Option 1, the Basic Multilingual Plane of ISO/IEC 10646-1 (commonly known as UNICODE) shall be used for the value of character entities in *extra.ents*. As exemplified in Figure 11, a character entity definition includes a mnemonic entity name, an entity value of the form `&#xHHHH;` (where HHHH consists of the four hex digits comprising the UNICODE number for the character), and an optional comment to clarify the meaning of the mnemonic.

Sometimes it is desirable to avoid processing the *extra.ents* entity during conversions. An alternative procedure for specifying character entities is to use numeric character references as defined in ISO 8879. Numeric character references need not be declared in the DTD and can be automatically generated from a UNICODE number. Descriptions of numeric character references can be placed in a `<refObjectList>` where each `<refObject>` has an id consisting of a UNICODE number followed by a value of the *lang* attribute.

8.3 MARTIF header

The structure of the header is defined in the framework (see Section 2, Figure 9). The header provides for a standard way in which the origin of the terminology file shall be identified and in which comments about it shall be recorded that will be helpful to a terminologist or translator using the document later. The header is not necessarily processed automatically, except perhaps to format it for presentation to a human reader.

The framework DTD declares that the header shall consist of three top-level elements: a required file description, an optional encoding description, and an optional revision description. The file description (`<fileDesc>`) shall consist of an optional title statement (comprising the actual title followed by an indication of the person responsible for the file), an optional publication statement (comprising a series of paragraphs describing where this document was published), and a required source description (consisting of a series of paragraphs describing where the information in this document originated).

After the file description comes an optional encoding description (`<encodingDesc>`), which can consist of a series of paragraphs describing the coding conventions used in the document. Specifically, in any case where a writing system (designated by a value of the *lang* attribute) uses any method of encoding involving non-ISO-646 characters other than the character entities from ISO 8879:1986, Annex D or ISO/IEC 10646-1 entities, then that method shall be documented in this element.

An optional revision description (`<revisionDesc>`) can follow the encoding description (if there is any revision information to report). If used, it shall consist of a series of *change* elements, each of which comprises a series of paragraphs. Each *change* element shall describe a change or set of changes that has been made in the document.

8.4 MARTIF text

8.4.1 MARTIF Components

As illustrated in Figures 5 and 6 (see 8.1), the text of a MARTIF document shall consist of an optional <front>, the <body>, and an optional <back>.

8.4.2 MARTIF front

Example 10 (see 8.1) does not include an optional front element. If the front element does occur in a MARTIF document, it shall contain explanatory prose that provides background information concerning the terminological entries. Wherever possible, the terminological entries contained in the body should be understandable without reference to information in the front element, even if this requires duplication of certain information throughout the body. For example, each terminological entry should contain an indication of subject field (domain) information rather than expecting the user to obtain this information from the front matter. Thus, the terminological entries will be more self-contained and more easily merged with other entries after interchange.

8.4.3 MARTIF body

As previously mentioned, the content of the mtf-body entity (shown in Figure 10, 8.5) declares the structure that shall be followed by the terminological entries. The DTD fragment for the body begins with the following SGML statement:

```
<!ELEMENT body - - (termEntry+)>.
```

This statement defines the body of the document by declaring that it shall consist of one or more <termEntry> elements. The following commentary on the structure of a <termEntry> is an informal re-statement of the information given formally in the statements of the body DTD component.

A terminological entry (<termEntry>) shall consist of optional auxiliary elements (contained in the parameter entity *termAux*) followed by one or more terminological information groups. As noted in Clause 4, one <termEntry> should ideally document one concept within a specified subject field, and each term representing the concept shall be documented in a terminological information group (<tig> or <ntig>). Multiple <tig>s or <ntig>s in the same language shall be contained in <langSet>.

A <tig> shall be used when there are no explicit subgroupings within the terminological information group. For example, a <tig> can consist of a term, grammatical information (classified as <termNote>), a definition, a bibliographic reference, and an indication of responsibility for maintenance of the <tig>. In such a case, the term, grammar, definition, reference, and responsibility all reside on the same logical level as elements of the <tig>.

An <ntig> shall be used when it is desirable to represent a secondary level of grouping within the terminological information group. For example, it can occur that a note or reference applies only to a <descrip> or <admin> element and not to the entire terminological information

group. In such a case, a Group element can be used to group the note or reference element explicitly with the appropriate element.

Example 5 and Examples 6 and 7 (see 5.2.4) show a sample <tig> vs. an <ntig>, respectively.

When <ntig> is used, <termGrp> shall be used to group information about the term itself. Any number of <descripGrp> elements can be used, each of which contains one descriptive element and any pointers, references, and notes that apply to it, and any number of <adminGrp> elements can be used, each of which contains one administrative element and any pointers, references, and notes that apply to it.

The content of the <term>, <descrip>, <admin>, and <note> elements allow various degrees of internal structure for the representation of highlighted elements and foreign words. The content of <descrip> is the most flexible, since contextual examples are placed in <descrip> elements. However, <ptr>, <ref>, and <note> elements shall not be used in the content of <term>, <descrip>, and <admin>.

In the MARTIF DTD, the basic method for showing relatedness is to use explicit embedding at several levels within an entry:

- Elements shall be embedded within an entry (but before any terminological information group) if they apply to the entire entry.
- All terminological information groups related to essentially the same concept shall be included in a single <termEntry>. In the event of multiple terminological information groups in a given language, language sets shall be embedded within the entry, with the <tig>s and <ntig>s embedded within the language sets to ensure that information related to the same language remains together within the <termEntry>.
- <term> elements, <termNote> elements, <descrip> elements, and <admin> elements shall be embedded within terminological information groups in such a way as to illustrate that they are related to each other.
- If needed, notes, pointers, and references shall be nested within <termGrp>, <termNoteGrp>, <adminGrp>, and <descripGrp> elements to show that they are related to a <term>, <termNote>, <admin>, or <descrip> element but not to the entire terminological information group.

Pointers, references, and cross-references (<ptr>, <ref>, <xref>) shall be used when it is necessary to show relatedness to other term entries, other parts of the document in question, or, in the case of <xref>, to external documents or files (see 5.2.5).

In some cases, however, there is need for a further level of reference in addition to those accommodated by MARTIF's embedding features. For instance, in some unusual situations, there may be a note on a note, which actually happens in the German <tig> included in Example 7 (5.2.4). In fact, in this particular case, there are *two* additional levels of embedding, i.e., a note on a note on a note. Inasmuch as such instances are rare, it is not desirable to implement additional layers of embedding in the MARTIF DTD. As shown in the example, relatedness is expressed in the following manner:

- The note shall be identified with a unique *id* and embedded in the appropriate <...Grp> element.

- A `<ref>` element embedded in the same `<...Grp>` element shall be used to point to the note in question. The actual content of the `<ref>` element shall be the content of the secondary note. In this kind of situation, the `<ref>` element functions as a `<ptr>` showing relatedness, but it is a pointer with a note. This mechanism can be used to account for multiple layers of relatedness without implementing complex mechanisms that would be rarely used or running the risk of creating recursive loops, which could be the case if the DTD were to allow `<note>` to be embedded inside a `<note>` itself.

The `%AuxInfo` entity consists of a sequence of any number of descriptive elements, pointers, references, notes, and administrative elements. This auxiliary information can appear in a `<termEntry>` before the first `<tig>`, in an `<ntig>` before the `<termGrp>`, or in a `<tig>` after the `<term>` and `<termNote>` elements. Auxiliary information applies to the entire enclosing element (the `<termEntry>` or `<ntig>`), rather than to the particular enclosed element following it. Thus, for example, a note can appear in several positions: as the first element of a `<termEntry>` (when it shall apply to the whole `<termEntry>`), as the first element of an `<ntig>` or any main element of a `<tig>` (when it shall apply to an entire terminological information group), as part of the content of a `<descrip>` (when it shall apply only to the `<descrip>`), or as part of a `<termGrp>`, `<termNoteGrp>`, `<descripGrp>`, or `<adminGrp>` (when it shall apply only to that group).

The framework and body components of the MARTIF DTD indicate what shall be the content of the various elements of a terminological entry. For example, the framework component defines the entities `bText` (basic text), `nText` (note text), and `dText` (descriptive text), which shall be used in the body.

8.4.4 MARTIF back

8.4.4.1 General requirements

The back element in general SGML documents can accommodate such common document features as indexes and annexes. In MARTIF, bibliographic entries and cross-references (with the exception of those between individual terminological entries) should usually be placed in the back element. Cross-references can include, among other items, references to responsibility identifiers, to nontextual illustrations as described in A.5.5 of Annex A, World Wide Web namespace identifiers (URLs and FPIs), or references to frequently cited text chunks and frequently referenced information such as concept system layouts.

Different types of information included in the back matter should be sorted according to type and listed together in `<refObjectList>`s identified according to the value of the attribute *type*. Annex B of this International Standard provides information on the representation of bibliographical information. Other types of information shall be treated as follows:

Each type of back-matter information shall be introduced by the GI `<refObjectList>`, which can be differentiated using the *type*= attribute, e.g., `<refObjectList type= 'URL'>`, `<refObjectList type='video'>`, etc., (see Annex A, A.11.4.1 for additional options). Each individual entry within a `<refObjectList>` shall be introduced by the GI `<refObject>`, which inherits its type from the `<refObjectList>` in which it appears. Each individual data item included in a `<refObject>` shall be introduced by the `<item>` GI. A `<refObject>` can include an unlimited number of `<item>` elements. The `<itemGrp>` element shall be used to cluster a given `<item>` with related `<note>`, `<ptr>`, and `<ref>` elements.

8.4.4.2 Bibliographic references

As noted in 5.2.4, there are three different ways to treat bibliographic information in a MARTIF document.

- All information occurs in the <termEntry> itself. This procedure is followed in some databases, but it poses the risk of redundancy, which in turn is a source of error and cost as a result of increased effort devoted to information management.
- If the full bibliographic entry is included in the back element, <ptr> or <ref> elements shall be used to link the term entry with the respective material in the back matter. The entire block of bibliographic information shall be introduced by <refObjectList type='bibl'>, and each bibliographic entry shall be introduced by the <refObject> GI. The individual bibliographic data categories shall be encoded using the <item> or the <itemGrp> GIs together with the attribute *type*. These data categories correspond to data categories specified in ISO 12083. The <itemSet> GI is used where ISO 12083 embeds multiple items inside other items. Example E.1 in Annex E, lines 131-132, 138-142, and 152-156, contain samples of this type of encoding. The GI <itemGrp> can also be used to contain a single <item> plus a <note>, <ptr>, or <ref>.
- If bibliographic information is contained in an external document or documents that are encoded in conformance with the DTD specified in ISO 12083, the <ptr> and <ref> elements shall be used to link the term entry with <xref> references included in the back element. These elements occur inside <itemGrp> elements as illustrated in Example 10 (see 8.1). They point to the respective external document(s).

8.4.4.3 Cross references

If a long contextual example or other piece of data is used several times in the body, it can be placed in the back element and referenced with a <ptr> or <ref> element from many different terminological information groups in order to avoid unnecessary duplication of data. This option applies particularly for the following types of data:

- Nontextual illustrations (see 8.4.4.4)
- External references from textual material (see 8.4.4.5)
- Namespace identifiers (see 8.4.4.6)

8.4.4.4 Nontextual illustrations

The back matter can also contain <xref> elements that point to other external files or documents, such as bit-mapped graphical images or other foreign, nontext, nonSGML data meant to be viewed with specialized software. Such materials can be linked to an entry in the body by using a <ref> element that points to an <xref> in the back, which in turn points to the external file, or they can be “bundle” and included in the back matter. This means that the data have been encoded to consist entirely of ISO 646:1983 characters and placed in an item in <back>. Of course, several entries can refer to the same foreign illustrative data item. Such items can include, for instance, spreadsheets, other databases, statistical models, virtual reality files, flight simulations, etc. Reference to these items can be entered at either the <tig> (<ntig>) or <termEntry> level.

8.4.4.5 External references from textual material

Sometimes a term entry can include an entailed term that references an external glossary where the term is defined. In such a case, the <hi> GI can be used as a pointer to an <xref> in the back matter.

EXAMPLE 11: Back-matter representation

```
1 ...
2 <descripGrp>
3   <descrip type='explanation'> This is a <hi type='entailed Term'
4     target='protein'>protein</hi> molecule.</descrip>
5   <ref type='sourceIdentifier' target='GDB-protein'>protein</ref>
6 </descripGrp>
7 ...
8 <back>
9
10  <refObjectList type='relatedTermbase'>
11    <refObject id='GDB-proteinLinks'>
12      <itemGrp>
13        <item id='protein'><xref target='gdbdir/protein.sgm'></xref></item>
14      </itemGrp>
15      <itemGrp>
16        <item id='amino_acid'>
17          <xref target='amino.acid.sgm'></xref>
18        </item>
19      </itemGrp>
20 ... [other links to entries in 'GDB-protein']
21    </refObject>
22  </refObjectList>
23
24  <refObjectList type='bibl'>
25    <refObject id='GDB-protein'>
26      <item type='title'>GDB-protein Glossary</item>
27      <item type='author'>Mike P. Biologist</item>
28      <item type='publisher'> etc. </item>
29    </refObject>
30  </refObjectList>
31 </back>
```

8.4.4.6 Namespace identifiers

Namespace identifiers are used to retrieve resources in the World Wide Web environment. Uniform Resource Locators (URLs) are unique character strings used to address files and servers. Formal Public Identifiers (FPIs) are unique identifiers for a representative of a given document. Both these identifiers constitute a kind of special bibliographical citation and serve as active cross-references. They have the potential to act in very different ways from normal citations, however, especially if the source database was configured for HTML and the URLs or FPIs are active links. The reference to a namespace identifier appears in a term entry in the body and contains the *anchor* for the identifier, which usually acts as a hypertext button in its source database. The actual links associated with the identifiers shall be compiled in the back matter using <refObjectList> with either the type='namespaceID' (for mixed lists), 'URL', or 'FPI'.

Example 12: Namespace identifiers

```

1 <termEntry id='prot32098'>
2   <ntig lang='en'>
3     <termGrp><term>zipper</term>
4     <termNote type='partOfSpeech'>noun</termNote></termGrp>
5     <descripGrp><descrip type='definition'>super secondary structures of <hi type='URL'
6       target='url34234'>proteins</hi>that look like zippers and can unzip and reform and that
7       exhibit a type of coiled structure</descrip> </descripGrp>
8   </ntig>
9 </termEntry>
10
11 <back>
12   <refObjectList type='URL'>
13     <refObject id='url34234'>
14       <itemGrp><item>http://cc.oulu.fi/~aiivanaai/cc.html</item>
15       <note>cf. PPS Glossary at the University of Oulu, Finland</note>
16     </itemGrp>
17   </refObject>
18 </refObjectList>
19 </back>

```

8.4.4.7 Responsibility references

Terminology databases document responsibility for data collection and maintenance based on three primary data categories: the type of terminology management transaction involved (see Annex A, A.10.1), the date (Annex A, A.10.2.1), and the person responsible for the transaction (Annex A, A.10.2.2). Annex A provides options for documenting this information, depending on the way the original database models this data.

Many databases can include internal or external entries documenting the individual persons responsible for creating and maintaining the database. Responsibility references can consist of pointers to these entries. Other databases do not necessarily include such entries, but do use undeclared responsibility codes or references. For purposes of exchange, any existing responsibility references shall be compiled in a list and placed in the back matter. If the database contains no separate responsibility entries, but does contain responsibility notation, the export routine shall extract this information based on the existing responsibility data items, assemble them into an appropriately coded <refObjectList>, and deposit this information in the back matter. Obviously, a generated list cannot include any information other than that provided by the existing data.

EXAMPLE 13: Responsibility entry

```

1 <refObjectList type='person'>
2   <refObject id='sew'>
3     <item type='name'>Sarah Elaine Wiggins</item>
4     <item type='jobTitle'>chief terminologist</item>
5     <item type='authorizationCode'>system operator</item>
6   </refObject>
7   <refObject id='mpp'>
8     <item type='name'>Maria Petra Pisana</item>
9     <item type='jobTitle'>Italian terminologist</item>
10    <item type='authorizationCode'>user</item>
11  </refObject>
12 </refObjectList>

```

Given the existence of a responsibility list in the back matter, the three components of the responsibility notation can be accommodated as follows:

EXAMPLE 14: Responsibility references

```

1 <ntig lang=en>
2   <termGrp>
3     <term>MARTIF</term>
4     <date type='origination'>1995-11-28</date>
5     <ptr type='originator' target='sew'>
6   </termGrp>
7   <descripGrp>
8     <descrip type='definition'>[A definition of MARTIF]</descrip>
9     <date type='update'>1993-07-25</date>

```

```

10      <ptr type='updater' target='mpp'>
11      </descripGrp>
12 </ntig>

```

8.5 Validation

Before transmitting a MARTIF file, the file should be validated using an SGML parser or some other SGML-aware software. Before a file can be parsed, it shall be reconfigured as a complete SGML document.

The MARTIF DTD shown in Figures 9-11 is completely self-contained and has been used to successfully parse the sample MARTIF documents, i.e., Example 10 (see 8.1) and Example E.1 shown in Annex E. These three parts of the DTD could have been placed in one file, but they have been kept as three files in order to exemplify the modularity of the DTD. This facilitates user addition of character entities. It also separates the body, which defines the structure of concept-oriented terminological entries, called <termEntry>s in a MARTIF document, from the framework portions of the DTD, which might also apply to other kinds of documents that consist of a header, optional front matter, a body, and back matter, but with a different structure inside the body.

Figure 9—The MARTIF framework DTD fragment

“ISO 12200: 1999/DTD for MARTIF (framework)”

```

<!--
OVERALL ORGANIZATION OF COMPLETE MARTIF DTD (framework and body):

Section 1: Building-block elements for MARTIF (entities and elements)

Section 2: High-level elements, including Header

Section 3: The text (front, body, and back) (NOTE: the internal structure
of the body element is not given here, see section 5)

Section 4: attributes (elements with same attributes are grouped)

(Section 5): the MARTIF body, i.e., the content of the body element in the
text element, is given in a separate file (mtf-body.DTD)
-->

<!-- ===== Section 1: Building Blocks ===== -->

<!-- GLOBAL ENTITIES -->

<!-- Basic text (for restricted use, such as terms) -->
<!ENTITY % bText    '(#PCDATA | hi)*' >

<!-- Text for notes (no recursion of note, allow foreign text) -->
<!ENTITY % nText    '(#PCDATA | hi | foreign)*' >

<!-- Text for descrip elements (allow embedded segments) -->
<!ENTITY % dText    '(#PCDATA | hi | foreign | s )*' >

```

```

<!-- ===== GLOBAL ATTRIBUTES ===== -->

<!-- language is a two or three character code that designates
      a language as represented by a certain writing system -->

<!ENTITY % a.global 'id ID #IMPLIED
      lang CDATA #IMPLIED' >

<!ENTITY % ta.global '%a.global
      TYPE CDATA #IMPLIED' >

<!-- ===== character entity set ===== -->

<!-- contents may be modified by user according to languages used -->

%mtf-ents;

<!-- ===== LOW-LEVEL ELEMENTS ===== -->

<!-- Elements used inside stretches of text -->

<!ELEMENT hi - - (#PCDATA) >
<!ELEMENT foreign - - (%bText;) >
<!ELEMENT s - - (%nText;) >

<!-- Elements used throughout MARTIF -->

<!ELEMENT note - - (%nText;) >
<!ELEMENT date - - (#PCDATA) >

<!ELEMENT ref - - (#PCDATA) >
<!-- Note: ptr is the only empty element -->
<!ELEMENT ptr - o EMPTY >

<!-- Elements used only in header, front, and back elements, not body -->

<!-- Note: p starts a paragraph; it need not have an end tag -->
<!ELEMENT p - o (%dText;) >

<!ELEMENT name - - (%dText;) >

<!ELEMENT refObjectList - - ( (refObject)+ ) >
<!ELEMENT refObject - - ((itemSet | itemGrp | item
      | ptr | ref | xref | note | date)+ ) >
<!ELEMENT item - - (xref | %dText)* >
<!ELEMENT itemGrp - - (item, (ptr | ref | xref | note | date)*) >
<!ELEMENT itemSet - - ((itemGrp | item)+) >

<!-- Reference to an object outside the document (use is discouraged) -->
<!ELEMENT xref - - (#PCDATA) >

<!-- ===== Section 2: High Level Elements ===== -->

<!-- NOTE: Document (MARTIF) consists of header and text.
Text consists of optional front, required body (which is in
the case of MARTIF a sequence of entries), and optional back
(for bibliographical and graphics references) -->

<!ELEMENT martif - - (martifHeader, text) >

<!-- *** Header dtd component is embedded here *** -->

```