

13 Terminological Databases

Since its first publication, this chapter has been rendered obsolete in several respects, chiefly as a result of the publication of ISO 12200, and a variant of it (TBX) which has been recently adopted by LISA, the Localisation Industry Standard Association. Work is currently ongoing in the ISO community to define a generic platform for terminological markup (ISO CD 16642, TMF : Terminological Markup Framework), in the light of which it is anticipated that the recommendations of the present chapter will be substantially revised. Readers are cautioned in particular that the discussion below of 'nested' and 'flat' structures is now far removed from current practices in the terminological field. A major revision of this chapter is planned for the next edition of these Guidelines.

Terminological information generally resides in *terminology databases (TDBs)*, but these collections of data can also be viewed as documents. A document containing terminological data is made up of *terminological entries*. Typically, a terminological entry treats a single concept and contains information on the assignment of single or multi-word terms to this concept. Bilingual and multilingual terminological entries deal with harmonized or very closely related concepts in two or more languages that are treated as functional equivalents in the context of a specific domain or subdomain. Terminological data can take the form of terminological databases (TDBs) or can be used to print hardcopy terminological documents, such as terminological dictionaries, technical vocabularies, or thesauri.

The TEI description of terminological data was originally designed primarily as a *terminology interchange format (TIF)* to allow users of terminology databases to exchange database records. The exchange of database records is especially important in practice because the structure of terminological records varies considerably from TDB to TDB, reflecting differences of design and of user needs. Users of TDBs frequently need to interchange data in order to access expert information and to prevent the duplication of effort, but differences in software, hardware, and methodology complicate interchange. A universal interchange format is a crucial element in making interchange easier.

The tag set defined in this chapter may also be used to mark up documents for the purpose of printing *terminological dictionaries* and *vocabularies*, or exchanging them in electronic form. Printed terminological documents differ from terminological databases in that they are frequently divided into sections and subsections and include prose text in introductions, etc.

Because printed terminological dictionaries differ from terminological databases, problems may arise if one attempts to use the same electronic document both for printing and to exchange records among databases. A printed terminological dictionary may contain material not suitably encoded for introduction into database records. Domain and subdomain information may be implied by the arrangement of `<termEntry>`s rather than by explicit domain specifications within the individual entries.

Other interchange difficulties include differences between term entry styles used in prescriptive and descriptive terminology work and problems arising from differences in the degree of detail used to classify data elements in different databases. (The term *data element* is used by terminologists to refer to *the smallest defined individual items of information*, regardless of whether they are represented as markup elements or attributes, or as database fields or columns. That is the usage followed here.) Procedures for addressing these various problems are treated in more detail in another document, the *TEI / LISA / ISO - TIF — Terminology Interchange Format — A Tutorial* (1993).¹⁰²

13.1 The Terminological Entry

The basic unit of terminology management is the *terminological entry*. A terminological entry documents information pertaining to a concept and generally speaking contains at least one *term*. In addition to the term, various kinds of descriptive and administrative data are recorded concerning the term, the concept to which it is assigned, and relationships to other terms and concepts. Administrative information supports the management of the terminology database or document.

A sample terminological entry consists of a series of components like the following:

subject field appearance of materials
English term opacity

¹⁰² This document is reprinted in *TermNet News*, no 40, 1993, pp 5–64; copies are also available from Infoterm, z.Hd. Herrn Dr. Gerhard Budin, Heinestraße 38, Postfach No. 130, A-1021 Vienna, Austria.

grammatical information, part of speech, English term noun
definition, English term degree of obstruction to the transmission of visible light
bibliographical source, English term information ASTM Standard E284
responsibility for English term information ASTM Technical Committee E12
German term Opazität
grammatical information, part of speech, German term noun
grammatical information, gender, German term feminine
definition, German term Maß für die Lichtdurchsichtigkeit
bibliographical source, German term information HFdn1983-382
responsibility, German term information DIN Technical Committee for paper products
French term opacité
grammatical information, part of speech, French term noun
grammatical information, gender, French term feminine
definition, French term rapport du flux lumineux incident au flux lumineux transmis ou réfléchi par un noircissement photographique
bibliographical source, French term information HJdi1986
responsibility, French term information C.I.R.A.D.

13.2 Tags for Terminological Data

The following sections define elements for use in tagging terminological data. The elements and attributes listed are based on empirical studies. The studies indicated the use of a wide variety of different data element types (data categories or database field types), but this variety can be reduced to a relatively small set of elements and attributes expressing notions common to most, if not all, TDBs. Those elements and attributes are defined here. In addition, the global TEI attributes defined in section 3.5 *Global Attributes*, and the elements and attributes defined in chapter 6 *Elements Available in All TEI Documents*, can all be used in terminological applications.

When tagging terminological data, three elements constitute the set of *non-floating elements*: <term>, <otherForm>, and <descrip>. All other elements function as *floating elements*, including: <admin>, <note>, <gram>, <bibl>, <biblFull>, <date>, <table>, <formula>, <figure>, and the linking elements (<ptr>, <xptr>, <ref>, and <xref>). The rules for combining floating with non-floating elements are spelled out below in section 13.3.1 *Nested Term Entries*, and in section 13.3.2 *Flat Term Entries Using Rules of Adjacency*.

<term> contains a single-word, multi-word, or symbolic designation which is regarded as a technical term. Attributes include:

type classifies the term using some typology.

Values any string of characters; for serious terminological work, values should be taken from the dictionary of data element types specified in ISO WD 12 620.

<termEntry> contains a single complete 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 very closely related concepts comprising one or more terms in each language and their associated descriptive and administrative data. Attributes include:

type classifies the term entry using some typology, preferably the dictionary of data element types specified in ISO WD 12 620.

Values any string of characters; for serious terminological work, values should be taken from the dictionary of data element types specified in ISO WD 12 620.

<tig> within a <termEntry> element, contains information elements associated with a single term. Attributes include:

type classifies the <tig> using some typology, preferably the dictionary of data element types specified in ISO WD 12 620.

- Values* any string of characters; for serious terminological work, values should be taken from the dictionary of data element types specified in ISO WD 12 620.
- <otherForm>** contains an alternate designation for the concept treated by the term entry, such as a synonym. Attributes include:
type classifies the **<otherForm>** using some typology, preferably the dictionary of data element types specified in ISO WD 12 620.
Values any string of characters; for serious terminological work, values should be taken from the dictionary of data element types specified in ISO WD 12 620.
- <ofig>** within a **<tig>** element, contains information elements relating to a single **<otherForm>**. Attributes include:
type classifies the other-form information group according to some convenient typology, preferably the dictionary of data element types specified in ISO WD 12 620.
Values any string identifying a class of **<ofig>**
- <gram>** within an entry in a dictionary or a terminological data file, contains grammatical information relating to a term, word, or form. Attributes include:
type classifies the grammatical information given according to some convenient typology — in the case of terminological information, preferably the dictionary of data element types specified in ISO WD 12 620.
Sample values include:
 pos part of speech (any of the word classes to which a word may be assigned in a given language, based on form, meaning, or a combination of features, e.g. noun, verb, adjective, etc.)
 gen gender (formal classification by which nouns and pronouns, and often accompanying modifiers, are grouped and inflected, or changed in form, so as to control certain syntactic relationships)
 num number (e.g. singular, plural, dual, ...)
 animate animate or inanimate
 proper proper noun or common noun
- <descrip>** within a **<termEntry>** element, contains a definition, context or explanation used to explain or define the concept represented by a **<term>** or an **<otherForm>**. Attributes include:
type classifies the description using some convenient typology, preferably the dictionary of data element types specified in ISO WD 12 620.
Suggested values include:
 definition The description provides all the information needed to differentiate one concept from all other related concepts in the given domain.
- <admin>** within a **<termEntry>** element, contains administrative information pertaining to data management and documentation of the entry. Attributes include:
type identifies the administrative event or information using some typology, preferably the dictionary of data element types specified in ISO WD 12 620.
Suggested values include:
 responsibility The admin element identifies the agency or individual responsible for the data element or entry.
 created The admin element describes the creation of the data element or entry.
 updated The admin element describes the update or modification of the data element or entry.
 approved The admin element describes the final approval of the data element or entry.
 domain The element indicates the subject area to which a concept pertains.
 subdomain The element indicates the subdomain of the subject area to which the concept pertains.

As indicated, these elements all possess a type attribute, used to classify the generic elements so as

to match the classifications used by TDBs. The type attributes allow specific items of information not defined in the DTD to be tagged as one of the defined elements with an appropriate type value. The possible values of type thus constitute a sizable open list. However, the attribute values used in the examples shown in this chapter are all taken from those defined by ISO 12 620: 1999 (Computer applications in terminology — Data Categories).

The <ofig> and <otherForm> elements are not necessary if each potential <otherForm> element is recast as a term in its own <tig>. For example, a term could be placed in a <tig type="synonym">.

When the base tag set described in this chapter is used, the following attributes are added to the set of global attributes:

group indicates the group (term and related elements) to which this element should be associated by specifying a string matching the n attribute value on an appropriate element.

depend indicates the parent element to which this element should be associated by specifying a string matching the n attribute value on an appropriate element.

grpPtr indicates the group (term and related elements) to which this element should be associated by specifying its unique identifier, where this is available.

depPtr indicates the parent element to which this element should be associated by specifying its unique identifier, where this is available.

For discussion of the usage of these attributes, see below, section 13.3.2 *Flat Term Entries Using Rules of Adjacency*.

Among the TEI core elements, the following are most likely to be found necessary in encoding terminological data; for fuller descriptions see the appropriate sections in chapter 6 *Elements Available in All TEI Documents*. In the case of the <date> element, it should be noted that the ISO format (YYYY-MM-DD) is preferred for terminology entries.

<note> contains a note or annotation. Attributes include:

type describes the type of note.

Values Values can be taken from any convenient typology of annotation suitable to the work in hand; e.g. annotation, gloss, citation, digression, preliminary, temporary

resp (responsible) indicates who is responsible for the annotation: author, editor, translator, etc.

Sample values include:

auth[or] note originated with the author of the text.

ed[itor] note added by the editor of the text.

comp[iler] note added by the compiler of a collection.

tr[anslator] note added by the translator of a text.

transcr[iber] note added by the transcriber of a text into electronic form.

(initials) note added by the individual indicated by the initials.

place indicates where the note appears in the source text.

Sample values include:

foot note appears at foot of page.

end note appears at end of chapter or volume.

inline note appears as a marked paragraph in the body of the text.

left note appears in left margin.

right note appears in right margin.

interlinear note appears between lines of the text.

app[aratus] note appears in the apparatus at the foot of the page.

anchored indicates whether the copy text shows the exact place of reference for the note.

Legal values are:

yes copy text indicates the place of attachment for the note.

no copy text indicates no place of attachment for the note.

target indicates the point of attachment of a note, or the beginning of the span to which the note is attached.

- Values* reference to the ids of element(s) which begin at the location in question (e.g. the id of an <anchor> element).
- targetEnd** points to the end of the span to which the note is attached, if the note is not embedded in the text at that point.
Values reference to the id(s) of element(s) which *end* at the location(s) in question, or to an empty element at the point in question.
- <ref> defines a reference to another location in the current document, in terms of one or more identifiable elements, possibly modified by additional text or comment. Attributes include:
target specifies the destination of the reference by supplying the value of the id attribute on one or more other elements in the current document.
Values One or more valid identifiers, separated by white space.
- <ptr> defines a pointer to another location in the current document in terms of one or more identifiable elements. Attributes include:
target specifies the destination of the pointer by supplying the values used on the id attribute of one or more other elements in the current document
Values One or more valid identifiers, separated by white space.
- <xref> defines a reference to another location in the current document, or an external document, using an extended pointer notation, possibly modified by additional text or comment.
- <xptr> defines a pointer to another location in the current document or an external document.
- <date> contains a date in any format. Attributes include:
calendar indicates the system or calendar to which the date belongs.
Values Recommended values include: *Gregorian, Julian, Roman, Mosaic, Revolutionary, Islamic.*
value gives the value of the date in some standard form, usually yyyy-mm-dd.
Values Any string representing a date in standard format; recommended form is ISO 8601:2000 5.2.1.1 Complete representation, extended format (yyyy-mm-dd)
certainty indicates the degree of precision to be attributed to the date.
Values Any appropriate value, e.g. *ca., approx, after, before.*
- <bibl> contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.
- <biblStruct> contains a structured bibliographic citation, in which only bibliographic subelements appear and in a specified order.
- <biblFull> contains a fully-structured bibliographic citation, in which all components of the TEI file description are present.
- <table> contains text displayed in tabular form, in rows and columns. Attributes include:
rows indicates the number of rows in the table.
Values If no number is supplied, an application must calculate the number of rows.
cols indicates the number of columns in each row of the table.
Values If no number is supplied, an application must calculate the number of columns.
- <figure> indicates the location of a graphic, illustration, or figure. Attributes include:
entity names the external entity within which the graphic image of the figure is stored.
Values the name of an external unparsed entity declared elsewhere in the DTD.
- <formula> contains a mathematical or other formula. Attributes include:
notation supplies the name of a previously defined notation used for the content of the element.
Values The name of a formal notation previously declared in the document type declaration.

Like all other elements defined in the TEI DTDs, all elements in the base tag set for terminology possess the following global attributes:

lang indicates the language of the element content, usually using a two- or three-letter code from ISO 639.

- n gives a number (or other label) for an element, which is not necessarily unique within the document.
- id provides a unique identifier for the element bearing the ID value.

Using the tags defined here, the example given above in section 13.1 *The Terminological Entry* might be tagged thus:¹⁰³

```

<!-- Example 2a: Nested Term Entry -->
<termEntry>
  <admin type="domain">
    appearance of materials </admin>
  <tig lang="en">
    <term> opacity </term>
    <gram type="pos"> n </gram>
    <descrip type="definition"> degree of obstruction to the
transmission of visible light </descrip>
    <ptr type="bibliographic" target="astm.e284"/>
    <admin type="responsibility" resp="ASTM E12"/>
  </tig>
  <tig lang="de">
    <term> Opazit&auml;t </term>
    <gram type="pos"> n </gram>
    <gram type="gen"> f </gram>
    <descrip type="definition"> Ma&szlig; f&uuml;r die
Lichtdurchsichtigkeit </descrip>
    <ref type="bibliographic" target="hfdn1983"> p. 383 </ref>
    <admin type="responsibility" resp="DIN TC for paper products"/>
  </tig>
  <tig lang="fr">
    <term> opacit&eacute; </term>
    <gram type="pos"> n </gram>
    <gram type="gen"> f </gram>
    <descrip type="definition"> rapport du flux lumineux
incident au flux lumineux transmis ou r&eacute;fl&eacute;chi
par un noircissement photographique </descrip>
    <ptr type="bibliographic" target="hjdi1986"/>
    <admin type="responsibility" resp="C.I.R.A.D."/>
  </tig>
</termEntry>

```

Both the `<ptr type="bibliographic" target="ASTM.E284">` and `<ref type='bibliographic' target='HFDN1983'>` elements in the example indicate links to complete bibliographical entries included in the back matter element of the same document. 'HFdn1983' is a source reference code for a book, generated according to ISO/TC 37 WI 18, *Coding of Bibliographic References in Terminology Work and Terminography* (1991). Its full bibliographic record would be:

```

<!-- Example 2b: Full Bibliographic Entry -->
<biblFull>
  <titleStmt id="hfdn1983">
    <title> W&ouml;rterbuch technischer Begriffe mit 4300
Definitionen nach DIN </title>
    <editor> Henry G. Freeman </editor>
  </titleStmt>
  <editionStmt>
    <edition> III </edition>
  </editionStmt>
  <extent> 703 pp </extent>
  <publicationStmt>
    <publisher> Beuth Verlag GmbH </publisher>
    <pubPlace> Berlin and K&ouml;ln </pubPlace>
    <date> 1983 </date>
  </publicationStmt>
  <sourceDesc>

```

¹⁰³ In this example, as in the others, white space has been liberally used for the sake of legibility; in practice most actual encodings would use less white space.

```

    <p>Compiled for the standards of the DIN (Deutsches
      Institut f&uuml;r Normung).</p>
  </sourceDesc>
</biblFull>

```

Further examples, including alternate encodings of this term entry, are given below in section 13.3.2 *Flat Term Entries Using Rules of Adjacency*, and section 13.3.3 *Flat Term Entries Using Group and Depend Attributes*.

The formal definition of these elements depends on which style of markup is being used; for discussion of the two styles, see the following section, 13.3 *Basic Structure of the Terminological Entry*. For the formal declarations for the two styles, see sections 13.4.1 *DTD Fragment for Nested Style*, and 13.4.2 *DTD Fragment for Flat Style*.

13.3 Basic Structure of the Terminological Entry

A terminological entry is identified with the `<termEntry>` tag and contains one or more terms marked with the tag `<term>`, which may appear with associated elements. A single term and its associated elements (such as `<gram>`, `<descrip>`, `<admin>`) constitute a *term information group*, `<tig>`. A `<termEntry>` may be made up of one or more `<tig>`s.

There are two structural descriptions for `<termEntry>`s:

- *nested* `<termEntry>`s
- *flat* `<termEntry>`s

The nested structure is preferred, especially for interchange with unknown partners. The flat structure provides an option that can be used between interchange partners whose systems exhibit fairly similar structures. The flat structure may also be used as an intermediate form for systems making the transition to the nested format.

13.3.1 Nested Term Entries

A nested `<termEntry>` represents the hierarchical relationships implicit in the terminological entry by utilizing the following principles of embedding and adjacency.

- Rule of embedding in nested term entries: Elements that constitute a part of another element are embedded inside the parent element.
- Rules of adjacency in nested term entries:
 - N1** Any element that appears in a `<termEntry>` outside a `<tig>` applies to the entire `<termEntry>`.
 - N2** Any element that appears in a `<tig>` before the `<term>` element applies to the entire `<tig>`.
 - N3** Any floating element that appears after a non-floating element (i.e., after `<term>`, `<otherForm>` or `<descrip>`) and before the next non-floating element, refers to the immediately preceding non-floating element unless otherwise indicated using the `depend` attribute. (See section 13.3.3 *Flat Term Entries Using Group and Depend Attributes*, for a full discussion of the `depend` attribute.)

The conversion routine that creates the nested entry infers the language of the `<tig>` from the language of the `<term>`, a process that can be construed as ‘upward inheritance’ from `<term>` to `<tig>`. Standard TEI ‘downward inheritance’ applies for all the elements embedded in the `<tig>`: their language is that of the `<tig>`, unless this default value is overridden by stating a new value.

An example of a nested term entry was given in section 13.2 *Tags for Terminological Data*.

13.3.2 Flat Term Entries Using Rules of Adjacency

The flat terminological entry does not use the <tig> element to enclose a term and its associated elements. Instead, it provides other mechanisms to express the relationships that occur within and among entries in a TDB, while at the same time allowing the different types of entries found in different source TDBs to be represented in very natural ways. The difference between the nested and flat terminological entries is that, while both can express the same information, the nested structure represents the logical hierarchy implicit within the entry by embedding elements in one another, while the flat entry does not represent the logical hierarchy within the entry in this way. Since many existing TDBs do not overtly indicate any hierarchical structure such as that represented in a nested entry, the flat entry may be more apt to reflect the organization of data elements within an entry found in the particular source TDB, whereas the nested entry more obviously characterizes an ideal abstract structure of the term entry. In flat entries, terms and their associated elements are grouped by means of the following rules of adjacency:

Rules of adjacency in flat termEntry elements

- F1** Any element that appears in a <termEntry> before the first <term> is assumed to apply to the entire <termEntry>.
- F2** Any floating element that appears after a non-floating element (i.e., after <term>, <other-Form> or <descrip>) and before the next non-floating element refers to the immediately preceding non-floating element unless otherwise indicated using the depend attribute. (See section 13.3.3 *Flat Term Entries Using Group and Depend Attributes*, for a full discussion of the depend attribute.)

Encoded using the flat style, the example given in section 13.2 *Tags for Terminological Data*, might look like this:

```

<!-- Example 3: Flat <termEntry>      -->
<termEntry>
  <admin type='domain'> appearance of materials </admin>
  <term lang='en'> opacity </term>
  <gram type='pos'> n </gram>
  <descrip type='definition'> degree of obstruction to the
  transmission of visible light </descrip>
  <ptr type='bibliographic' target='ASTM.E284' />
  <admin type='responsibility' resp='ASTM E12'></admin>
  <term lang='de'> Opazit&uuml;t </term>
  <gram type='pos'> n </gram>
  <gram type='gen'> f </gram>
  <descrip type='definition'> Ma&szlig; f&uuml;r die
  Lichtdurchsichtigkeit
  </descrip>
  <ref type='bibliographic' target='HFDN1983'> p. 383 </ref>
  <admin type='responsibility' resp='DIN TC for paper products'>
  </admin>
  <term lang='fr'> opacit&eacute; </term>
  <gram type='pos'> n </gram>
  <gram type='gen'> f </gram>
  <descrip type='definition'> rapport du flux lumineux
  incident au flux lumineux transmis ou r&eacute;flect&eacute;
  par un noircissement photographique </descrip>
  <ptr type='bibliographic' target='HJDI1986' />
  <admin type='responsibility' resp='C.I.R.A.D.'> </admin>
</termEntry>

```

13.3.3 Flat Term Entries Using Group and Depend Attributes

In practice, there are term entries where elements are ordered in such a way that the rules of adjacency cannot be used. For instance, in Example 3 the <ptr> and <ref> linking elements refer to the immediately preceding <descrip> information. The <admin type='responsibility'> elements as represented here also refer to the <descrip> element. It may, however, be desirable for the bibliographic reference to refer not only to the quoted material in the descriptive element, but also to the term itself. Because the second rule of adjacency dictates that all floating elements following a non-floating element

refer to that non-floating element, a mechanism is required to ‘point’ to the <term> if the floating element depends on the <term> itself.

There are also other exceptions to the adjacency rules: in some term entries elements are associated with a <term> other than the immediately preceding <term>. Such entries may be called *discontiguous flat term entries*, since the constituents of a term information group may not be adjacent. In such entries, information pertaining to the entire terminological entry may not always appear at the beginning of the entry (i.e., prior to the introduction of a term).

Such an entry might be encoded as follows:

```

<!-- Example 4: Discontiguous Flat <termEntry> -->
<termEntry n='texyz'>
  <term lang='en' n='1'> opacity </term>
  <gram type='pos' depend='1'> n </gram>
  <term lang='de' n='2'> Opazit&auml;t </term>
  <gram type='pos' depend='2'> n </gram>
  <gram type='gen' depend='2'> f </gram>
  <term lang='fr' n='3'> opacit&eacute; </term>
  <gram type='pos' depend='3'> n </gram>
  <gram type='gen' depend='3'> f </gram>

  <descrip type='definition' group='1' n='endes1'> degree of
obstruction to the transmission of visible light </descrip>
  <descrip type='definition' group='2' n='dedes1'> Ma&szlig; f&uuml;r die
Lichtdurchsichtigkeit </descrip>
  <descrip type='definition' group='3' n='frdes1'> rapport du
flux lumineux incident au flux lumineux transmis ou
r&eacute;f&eacute;chi par un noircissement photographique
</descrip>
  <ptr type='bibliographic' depend='endes1' target='ASTM.E284'>/>
  <admin type='responsibility' depend='endes1' resp='ASTM E12'> </admin>
  <ref type='bibliographic' depend='dedes1' target='HFDN1983'>
p. 383 </ref>
  <admin type='responsibility' depend=dedes1
resp='DIN.TC.for.paper'></admin>
  <ptr depend='frdes1' type='bibliographic' target='HJDI1986'>/>
  <admin type='responsibility' depend=frdes1
resp='C.I.R.A.D.'> </admin>
  <admin type='domain' depend='texyz'> appearance of materials
</admin>
</termEntry>

```

In the above example, depend elements indicate that the material tagged with this attribute is related to the targeted element. The group elements indicate that the information so marked is part of an implicit <ti g>, i.e. that it pertains either to the term or to the entire implicit <ti g>. Items linked to other elements by depend do not require the group attribute because they are associated with the group already by virtue of their relation to elements that are themselves associated with the group.

So as to describe appropriate relationships in discontiguous flat <termEntry>s, it is necessary to define a pointing mechanism that allows any non-adjacent element to be related to an implicit term information group and therefore to the <term> with which it is associated or to some other specific element.

Two methods are provided to represent this association. For terminology files in which unique identifiers for all <term> elements cannot be assumed (as will often be the case in interchange), the group and depend attributes should be used. For terminology files in which unique identifiers can be provided, the grpPtr and depPtr attributes should be used. The two pairs of attributes have identical significance as far as the association of elements is concerned.

The group attribute associates an element with a specific term, or with an implicit term information group: its value must be the same as the n attribute on the <term> element being pointed to. During interchange, the group attribute would be used to extract and assemble all the elements related to a specific term information group from a discontiguous flat <termEntry> by matching them to the n attributes

on the terms. The group pointer accounts for the kind of relationship represented by the principle of embeddedness within a <ti g> in a nested term entry.

The depend attribute associates an element with some other specific element: its value must be the same as the n attribute on the element being pointed to. As shown in the last line of Example 4, the depend attribute can also point to the entire terminological entry by targeting a value of n indicated in the <termEntry> element. If for any reason the grammatical information pertaining to a term does not follow the term immediately, this information must be linked to the term with the depend attribute.

In terms of the extended pointer notation defined in chapter 14 *Linking, Segmentation, and Alignment*, the specification group="2" is synonymous with HERE ANCESTOR (1 TERMENTRY) DESCENDANT (1 TERM N 2), and the specification depend="3" is synonymous with HERE ANCESTOR (1 TERMENTRY) DESCENDANT (1 * N 3).

To summarize the behavior of group and depend, the group attribute identifies an implicit <ti g>, whereas the depend attribute implies relatedness. If there is any ambiguity with respect to the rules of adjacency, one should use depend.

In Example 4, the English term 'opacity' is identified as n="1", and all other elements associated with this <ti g> are marked as group="1"; in German, the term and all its associated elements are identified as n="2" and group="2", respectively; in French, the term and associated elements are marked group="3". Since the bibliographical references are displaced from the descriptive information with which they are associated, the descriptions are identified with n="endes1", n="dedes1", and n="frdes1", respectively. The <ptr> and <ref> elements are then identified with depend attributes that target the appropriate descriptions. Even if the elements in the entry were adjacent to each other in the entry, this convention would be essential if one wanted to indicate that the source applied to the <term> and hence to the entire <ti g>, rather than just to the <descrip> element itself.

13.3.4 References between Term Entries

Terminology documents utilize a variety of cross-references between <termEntry>s, for instance to link to bibliographic entries or between equivalents in different languages, synonyms and related terms and concepts. These references are usually implemented using the TEI linking elements <ptr> and <ref>, together with a value of the attribute type. If, as is the case with the reference to ASTM E284, the total bibliographic source description is contained in the 'target' element of the linking element, use <ptr>. If, on the other hand, a page number is included, this page number must appear as the content of a linking element introduced by the <ref> element.

Examples:

```
<ptr type="bibliographic" target="astm.e284"/>
```

or

```
<ref type="bibliographic" target="hfdn1983"> p. 383 </ref>
```

If the full bibliographical citation is included in the <termEntry> itself, linking elements are unnecessary and the citation can be marked using the <bibl>, <biblStruct>, or <biblFull> elements. For further discussion of bibliographic citations and references, see section 6.10 *Bibliographic Citations and References*.

13.4 Overall Structure of Terminological Documents

To enable the base tag set for terminology, a parameter entity TEI.terminology must be declared within the document type subset, the value of which is INCLUDE, as further described in section 3.3 *Invocation of the TEI DTD*. A document using this base tag set and no other additional tag sets will thus begin as follows:

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE TEI.2 PUBLIC "-//TEI P4//DTD Main Document Type//EN" "tei2.dtd" [
  <!ENTITY % TEI.XML 'INCLUDE' >
  <!ENTITY % TEI.terminology 'INCLUDE' >
]>
```

This declaration makes available all of the elements described in this chapter, in addition to the core elements described in chapter 6 *Elements Available in All TEI Documents*. The default structure for terminological documents is similar to that defined by chapter 7 *Default Text Structure*: within the <TEI.2> element they contain a <teiHeader> and a <text>. The <text> element, in turn, contains as usual a <body> element, optionally preceded by a <front> and followed by a <back>. The <body> may contain a series of <termEntry> elements, which may optionally be grouped into sections tagged with the same elements (<div>, <div0>, <div1>, etc.) as defined in section 7.1 *Divisions of the Body*.

<text> contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample.

<body> contains the whole body of a single unitary text, excluding any front or back matter.

<div> contains a subdivision of the front, body, or back of a text.

<div0> contains the largest possible subdivision of the body of a text.

<div1> contains a first-level subdivision of the front, body, or back of a text (the largest, if <div0> is not used, the second largest if it is).

<div2> contains a second-level subdivision of the front, body, or back of a text.

<div3> contains a third-level subdivision of the front, body, or back of a text.

<div4> contains a fourth-level subdivision of the front, body, or back of a text.

<div5> contains a fifth-level subdivision of the front, body, or back of a text.

<div6> contains a sixth-level subdivision of the front, body, or back of a text.

<div7> contains the smallest possible subdivision of the front, body or back of a text, larger than a paragraph.

In order to support both the flat and the nested styles of markup, three distinct DTD fragments for terminology are provided.

- teiterm2
- teite2n
- teite2f

In file teiterm2.dtd, the top-level elements for the terminology base are defined, and a subordinate parameter entity, termtags is defined and referred to. By default, this entity refers to file teite2n.dtd, which defines the DTD for nested markup; if the flat style of markup is to be used, the document's DTD subset should define termtags as referring to the file teite2f.dtd, as shown in the examples in section 13.3.2 *Flat Term Entries Using Rules of Adjacency*.

```

<!-- 13.4: TEIterm2.DTD: Base tag set for terminological data-->
<!--Text Encoding Initiative Consortium:
Guidelines for Electronic Text Encoding and Interchange.
Document TEI P4, 2002.
Copyright (c) 2002 TEI Consortium. Permission to copy in any form
is granted, provided this notice is included in all copies.
These materials may not be altered; modifications to these DTDs should
be performed only as specified by the Guidelines, for example in the
chapter entitled 'Modifying the TEI DTD'
These materials are subject to revision by the TEI Consortium. Current versions
are available from the Consortium website at http://www.tei-c.org-->
<!--First, embed the default text structure elements.-->
<![%TEI.singleBase;[
<!ENTITY % TEI.structure.dtd PUBLIC "-//TEI P4//ELEMENTS Default Text
Structure//EN" 'teistr2.dtd' >
%TEI.structure.dtd;
]]>
<!ENTITY % termtags PUBLIC "-//TEI P4//ELEMENTS Terminological Databases
(Nested)//EN" 'teite2n.dtd' >%termtags;
<!-- end of 13.4-->

```

In file teiterm2.ent, terminology-specific extensions to the TEI element class system are defined, including the classes terminology, comp.terminology, terminologyInclusions, and terminologyMisc.

```

<!-- 13.4: TEIterm2.ent: Base tag set for terminological data-->
<!--Text Encoding Initiative Consortium:

```

Guidelines for Electronic Text Encoding and Interchange.
 Document TEI P4, 2002.
 Copyright (c) 2002 TEI Consortium. Permission to copy in any form
 is granted, provided this notice is included in all copies.
 These materials may not be altered; modifications to these DTDs should
 be performed only as specified by the Guidelines, for example in the
 chapter entitled 'Modifying the TEI DTD'
 These materials are subject to revision by the TEI Consortium. Current versions
 are available from the Consortium website at <http://www.tei-c.org>-->

```

<!ENTITY % x.comp.terminology "" >
<!ENTITY % m.comp.terminology "%x.comp.terminology; %n.termEntry;">
<!ENTITY % seq '(%m.common; | %m.comp.terminology;)* ' >
<!ENTITY % mix.terminology '| %m.comp.terminology;' >
<!ENTITY % x.terminologyInclusions "" >
<!ENTITY % m.terminologyInclusions "%x.terminologyInclusions;
%n.date; | %n.dateStruct; | %n.note; | %n.ptr; | %n.ref; | %n.xptr; |
%n.xref;">
<!ENTITY % x.terminologyMisc "" >
<!ENTITY % m.terminologyMisc "%x.terminologyMisc; %n.admin; |
%n.descrip;">
<!--Add attributes to the set of global attributes:-->
<!ENTITY % a.terminology '
  group CDATA #IMPLIED
  grpPtr IDREF #IMPLIED
  depend CDATA #IMPLIED
  depPtr IDREF #IMPLIED'>
<!-- end of 13.4-->

```

13.4.1 DTD Fragment for Nested Style

In file teite2n.dtd the following definitions are found, which define the elements used in the nested markup style:

```

<!-- 13.4.1: Elements for nested-style terminological data-->
<!--The nested structure is used for data interchange and represents a
canonical structured form for terminology entries, which differs
from the less structured forms frequently used to store data in
terminological databases.-->
<!ELEMENT termEntry %om.R0;
((%m.terminologyMisc; | %m.terminologyInclusions; | %m.Incl;)*,
 (tig, (%m.Incl; | %m.terminologyInclusions;)*)*
>
<!ATTLIST termEntry
  %a.global;
  type CDATA #IMPLIED
  TEIform CDATA 'termEntry' >
<!--Notes, descrip(s) and admin(s) are allowed in the termEntry
to provide documentation that applies to the whole entry.-->
<!--tig='term information group'-->
<!--ofig='otherform information group'-->
<!ELEMENT tig %om.R0;
((%m.terminologyMisc; | %m.terminologyInclusions; | %m.Incl;)*,
 (term, (gram | %m.terminologyInclusions; | %m.Incl;)*),
 (%m.terminologyMisc;),
 (%m.terminologyInclusions; | %m.Incl;)*)*,
 (ofig, (%m.terminologyInclusions; | %m.Incl;)*)*
>
<!ATTLIST tig
  %a.global;
  type CDATA #IMPLIED
  TEIform CDATA 'tig' >
<!--Order is significant: term, descrip(s), ofig(s) or otherform(s)-->
<!ELEMENT ofig %om.R0;
((%m.terminologyMisc; | %m.Incl;)*, (otherForm, (gram | %m.Incl;)*),
 (%m.terminologyMisc;), (%m.Incl;)*)*>
<!ATTLIST ofig
  %a.global;

```

```

        type CDATA #IMPLIED
        TEIform CDATA 'ofig' >
<!ELEMENT otherForm %om.R0; %paraContent;>
<!ATTLIST otherForm
    %a.global;
    type CDATA #IMPLIED
    TEIform CDATA 'otherForm' >
<!ELEMENT descrip %om.R0; %paraContent;>
<!ATTLIST descrip
    %a.global;
    type CDATA #IMPLIED
    TEIform CDATA 'descrip' >
<!ELEMENT admin %om.R0; %paraContent;>
<!ATTLIST admin
    %a.global;
    type CDATA #IMPLIED
    date %ISO-date; #IMPLIED
    resp CDATA #IMPLIED
    TEIform CDATA 'admin' >
<!--We define a.dictionaries as the empty string,
since we are not now using the tag set for
dictionaries.-->
<!ENTITY % a.dictionaries ''>
<!ELEMENT gram %om.R0; %paraContent;>
<!ATTLIST gram
    %a.global;
    %a.dictionaries;
    type CDATA #IMPLIED
    TEIform CDATA 'gram' >
<!-- end of 13.4.1-->

```

13.4.2 DTD Fragment for Flat Style

In file teite2f.dtd the following definitions, which provide support for the flat markup style, are found:

```

<!-- 13.4.2: Elements for flat-style terminological data-->
<!--The flat structure is used to represent a variety of terminology
documents that occur in practice and which do not follow the form
of the nested interchange
format. The flat representation allows for a less
rigid structure, but provides a rich mechanism for reflecting
inter-element relations.-->
<!--The declaration of termEntry enforces appearance of at least one term
element in a termEntry, which may be preceded by descrip, admin, note,
otherform, or gram. There may be multiple notes, admins, descripts
otherforms, and grams appearing in any order. xRef, date, biblRef
can appear in all positions in termEntry.-->
<!ELEMENT termEntry %om.R0;
( (%m.terminologyMisc; | otherForm | gram |
%m.terminologyInclusions; | %m.Incl;)*, (term,
(%m.terminologyMisc; | otherForm | gram | %m.terminologyInclusions; |
%m.Incl;)* )+ )
>
<!ATTLIST termEntry
    %a.global;
    type CDATA #IMPLIED
    TEIform CDATA 'termEntry' >
<!ELEMENT otherForm %om.R0; %paraContent;>
<!ATTLIST otherForm
    %a.global;
    type CDATA #IMPLIED
    TEIform CDATA 'otherForm' >
<!ELEMENT descrip %om.R0; %paraContent;>
<!ATTLIST descrip
    %a.global;
    type CDATA #IMPLIED
    TEIform CDATA 'descrip' >

```

```

<!ELEMENT admin %om.R0; %paraContent;>
<!ATTLIST admin
  %a.global;
  type CDATA #IMPLIED
  date %ISO-date; #IMPLIED
  resp CDATA #IMPLIED
  TEIform CDATA 'admin' >
<!--We define a.dictionaries as the empty string,
since we are not now using the tag set for
dictionaries.-->
<!ENTITY % a.dictionaries ''>
<!ELEMENT gram %om.R0; %paraContent;>
<!ATTLIST gram
  %a.global;
  %a.dictionaries;
  type CDATA #IMPLIED
  TEIform CDATA 'gram' >
<!-- end of 13.4.2-->

```

13.5 Additional Examples of Term Entries

The tag set defined in this chapter is designed to accommodate the variety of structures that occur in TDBs; this section shows how the same information may be encoded in different ways, depending on local convenience or preferences. Example 5 gives an entry from an ISO terminological standard. Example 6 treats this English-French equivalent pair as a single nested terminological entry, whereas Example 7 splits the information into two nested entries with cross-references. Example 8 shows the same data as a flat terminological entry with adjacent elements, whereas Example 9 groups the elements according to element type, which requires the use of pointers in order to reconstruct the implicit terminological information group from discontinuous elements.

13.5.1 Example Term Entry from ISO 472

The following term entry is taken from ISO 472:1988, *Plastics — Vocabulary*, Bilingual edition (Geneva: ISO, 1988), p. 84. The original uses typographic characteristics to represent different data element types within the term entry, not all of which have been retained in the reproduction of this sample. As prescribed by ISO layout guidelines,¹⁰⁴ the original text is printed in Helvetica, with English and French information presented in two parallel columns; head terms appear in bold face, notes in a smaller font size than the main text, and terms referred to in the cross references are printed in italics.

thermal degradation The entirety of all deleterious chemical modifications of plastic at elevated temperature.

NOTE — It is essential to report the temperature and other environmental conditions at which the phenomenon is studied.

See also *ageing*, *degradation* and *deterioration*.

décomposition thermique Ensemble de toutes les modifications chimiques nuisibles d'un plastique à température élevée.

NOTE — Il est essentiel d'indiquer la température et les autres conditions d'environnement dans lesquelles le phénomène est étudié.

Voir aussi *viellissement*, *dégradation* et *détérioration*.

13.5.2 The Example Treated as a Single Term Entry in Nested Form

This treatment assumes that both the English and French terms are treated together in the same entry. The elements grouped together at the top of the term entry apply to the entire entry. Only the first of the three cross-referenced terms is included in this example; it is represented by a <pt r> link which targets a term entry (related concept) contained in the same document. The id values used here are purely arbitrary.

```

<termEntry id="te84.11">
  <admin type="domain"> plastics </admin>

```

¹⁰⁴ ISO 10241, *Preparation and layout of international terminology standards*, 1993.

```

<ref type="bibliographic" target="iso.472-1988"> p. 84 </ref>
<admin type="creation" date="1988" resp="ISO/TC 61, Plastics"/>
<ptr type="relatedTerm" target="te04.06"/>
<tig lang="en">
  <term> thermal degradation </term>
  <gram type="pos"> n </gram>
  <descrip type="definition"> The entirety of all
    deleterious chemical modifications of plastic at
    elevated temperature. </descrip>
  <note> It is essential to report the
    temperature and other environmental conditions at which
    the phenomenon is studied. </note>
</tig>
<tig lang="fr">
  <term> d&eacute;composition thermique </term>
  <gram type="pos"> n </gram>
  <gram type="gen"> f </gram>
  <descrip type="definition"> Ensemble de toutes les
    modifications chimiques nuisibles d'un plastique &agrave;
    temp&eacute;rature &eacute;lev&eacute;e. </descrip>
  <note> Il est essentiel d'indiquer la
    temp&eacute;rature et les autres conditions d'environnement
    dans lesquelles le ph&eacute;nom&egrave;ne est
    &eacute;tudi&eacute;. </note>
</tig>
</termEntry>
<!-- Referenced term entry: -->
<termEntry id="te04.06">
  <tig lang="en">
    <term> ageing </term><!-- ... -->
  </tig>
  <tig lang="fr">
    <term> vieillissement </term><!-- ... -->
  </tig>
</termEntry>

```

13.5.3 The Example Treated as Two Separate Term Entries in Nested Form

This example takes cognizance of the fact that some TDBs treat each term in a single <termEntry> instead of grouping all the information for a single concept into a single <termEntry>. The rationale behind this approach is frequently that no two languages truly provide harmonized concepts, although in the case of standardized terminology it can generally be assumed that concepts have been harmonized. The significant difference in encoding that occurs in this type of system is that <ptr> linking elements are required more frequently to link to term equivalents and related terms in other entries in the same document. Since there is only one <tig> in each entry, the <ptr> element could come at the beginning, as shown in the previous example, or inside the <tig> as shown below.

```

<termEntry id="te84.11.en">
  <admin type="domain"> plastics </admin>
  <ref type="bibliographic" target="iso.472-1988"> p. 84 </ref>
  <admin type="creation" date="1988" resp="ISO/TC 61, Plastics"/>
  <tig lang="en">
    <term> thermal degradation </term>
    <gram type="pos"> n </gram>
    <descrip type="definition"> The entirety of all
      deleterious chemical modifications of plastic at
      elevated temperature. </descrip>
    <note> It is essential to report the
      temperature and other environmental conditions
      at which the phenomenon is studied. </note>
    <ptr type="relatedTerm" target="te04.06.en"/>
    <ptr lang="fr" type="equivalent" target="te84.11.fr"/>
  </tig>
</termEntry>
<termEntry id="te84.11.fr">

```

```

<admin type="domain"> plastics </admin>
<ref type="bibliographic" target="iso.472-1988"> p. 84 </ref>
<admin type="creation" date="1988" resp="ISO/TC 61, Plastics"/>
<tig lang="fr">
  <term> d&eacute;composition thermique </term>
  <gram type="pos"> n </gram>
  <gram type="gen"> f </gram>
  <descrip type="definition"> Ensemble de toutes les
    modifications chimiques nuisibles d'un plastique &agrave;
    temp&eacute;rature &eacute;lev&eacute;e. </descrip>
  <note> Il est essentiel d'indiquer
    la temp&eacute;rature et les autres conditions
    d'environnement dans lesquelles le ph&eacute;nom&egrave;ne
    est &eacute;tudi&eacute;. </note>
  <ptr type="relatedTerm" target="te04.06.fr"/>
  <ptr lang="en" type="equivalent" target="te84.11.en"/>
</tig>
</termEntry>
<!-- Referenced term entry: -->
<termEntry id="te04.06.en">
  <tig lang="en">
    <term> ageing </term><!-- ... -->
  </tig>
</termEntry>
<termEntry id="te04.06.fr">
  <tig lang="fr">
    <term> vieillissement </term><!-- ... -->
  </tig>
</termEntry>

```

13.5.4 The Example Treated as a Flat Term Entry Using Adjacency Rules

This version of Example 5 uses a flat style of encoding, following the pattern of many existing TDBs; elements associated with a given term follow it immediately:

```

<termEntry id='TE84.11'>
  <admin type='domain'> plastics </admin>
  <ref type='bibliographic' target='ISO.472-1988'> p. 84 </ref>
  <admin type='creation' date='1988'
    resp='ISO/TC 61, Plastics'> </admin>
  <term lang='en'> thermal degradation </term>
  <gram type='pos'> n </gram>
  <descrip type='definition'> The entirety of all deleterious
    chemical modifications of plastic at elevated temperature.
  </descrip>
  <note> It is essential to report the temperature and other
    environmental conditions at which the phenomenon is studied.
  </note>
  <term lang='fr'> d&eacute;composition thermique </term>
  <gram type='pos'> n </gram>
  <gram type='gen'> f </gram>
  <descrip type='definition'> Ensemble de toutes les
    modifications chimiques nuisibles d'un plastique &agrave;
    temp&eacute;rature &eacute;lev&eacute;e. </descrip>
  <note> Il est essentiel d'indiquer la temp&eacute;rature et les
    autres conditions d'environnement dans lesquelles le
    ph&eacute;nom&egrave;ne est &eacute;tudi&eacute;. </note>
  <ptr type='relatedTerm' target='TE04.06'>
</termEntry>

<!-- Referenced term entry: -->

<termEntry id='TE04.06'>
  <term lang='en'> ageing </term>
  <!-- ... -->
  <term lang='fr'> vieillissement </term>

```



```

    <!-- ... -->
  </termEntry>

```

13.5.5 The Example Treated as a Flat Term Entry Not Using Adjacency Rules

Many translation-oriented terminologists who work with half-screen popup windows prefer the following layout because it enables them to see the various <term> options at the top part of their display window without having to scroll into the body of the <termEntry>. Note in this case that the <ref> element links the bibliographic information to the entire entry.

```

<termEntry id='TE84.11' n='te84.11'>
  <term lang='en' n='1'> thermal degradation </term>
  <gram type='pos' depend='1'> n </gram>
  <term lang='fr' n='2'> d&eacute;composition thermique
  </term>
  <gram type='pos' depend='2'> n </gram>
  <gram type='gen' depend='2'> f </gram>

  <descrip type='definition' group='1'> The entirety of all
  deleterious chemical modifications of plastic at elevated
  temperature. </descrip>
  <descrip type='definition' group='2'> Ensemble de toutes les
  modifications chimiques nuisibles d'un plastique &agrave;
  temp&eacute;rature &eacute;lev&eacute;e. </descrip>
  <note group='1'> It is essential to report the temperature and
  other environmental conditions at which the phenomenon is
  studied. </note>
  <note group='2'> Il est essentiel d'indiquer la temp&eacute;rature et
  les autres conditions d'environnement dans lesquelles le
  ph&eacute;nom&egrave;ne est &eacute;tudi&eacute;. </note>

  <ptr type='relatedConcept' target='TE04.06' />
  <admin depend='te84.11' type='domain'> plastics </admin>
  <ref type='bibliographic' depend='te84.11' target='ISO.472-1988'>
  p. 84 </ref>
  <admin depend='te84.11' type='creation' date='1988'
  resp='ISO/TC 61, Plastics'>
  </admin>
</termEntry>
<!-- Referenced term entry: -->
<termEntry id='TE04.06' n='te04.06'>
  <term lang='en' n='1'> ageing </term>
  <!-- ... -->
  <term lang='fr' n='2'> vieillissement </term>
  <!-- ... -->
</termEntry>

```


IV: Additional Tag Sets

