Encoding for Interchange: An Introduction to the TEI

Lou Burnard   C. M. Sperberg-McQueen

Draft version of November 21, 1994

Contents

Abstract

The purpose of this document is to provide a brief introduction to the recommendations of the Text Encoding Initiative (TEI). It shows how these recommendations may be used to encode a wide variety of commonly encountered textual features, in such a way as to maximize the usability of electronic transcriptions and to facilitate their interchange among scholars using different computer systems.

This tutorial discusses the basic principles of encoding texts, and describes most of the TEI "core" tag set and most of the elements defined in the TEI "base tag set for prose". It does not address other more specialized tag sets. However, the elements and attributes described here should be adequate for the encoding of a wide variety of different kinds of material to a reasonable degree of detail. Some basic knowledge of SGML is assumed.

1 Introduction

The Text Encoding Initiative (TEI) Guidelines are addressed to anyone who wants to interchange information stored in an electronic form. They emphasize the interchange of textual information, since this is the form of information which is most frequently interchanged within the scholarly community, but other forms of information such as images and sound are also addressed. The Guidelines are equally applicable in the creation of new resources, and the interchange of existing ones.

The Guidelines provide a means of making explicit certain features of a text in such a way as to to aid the processing of that text by computer programs running on different machines. This process of making explicit we call "markup" or *encoding*. Any textual representation on a computer uses some form of markup; the TEI came into being partly because of the enormous variety of mutually incomprehensible encoding schemes currently besetting scholarship, and partly because of the expanding range of scholarly uses now being identified for texts in electronic form.

The TEI Guidelines use the *Standard Generalized Markup Language* (SGML) to define its encoding scheme. SGML is an international standard (ISO 8879), used increasingly throughout the information processing industries, which makes possible a formal definition of an encoding scheme, in terms of *elements* and *attributes*, and rules governing their appearance within a text. The TEI's use of SGML is ambitious in its complexity and generality, but it is fundamentally no different from that of any other SGML mark up scheme, and so any general purpose SGML-aware software is able to process TEI-conformant texts.

The TEI is sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing. Funding has been provided in part from the U.S. National Endowment for the Humanities, Directorate XIII of the Commission of the European Communities, the Andrew W. Mellon Foundation, and the Social Science and Humanities Research Council of Canada. Its Guidelines were published in May 1994, after six years of development involving many hundreds of scholars from different academic disciplines worldwide.

At the outset of its work, the overall goals of the TEI were defined by the closing statement of a planning conference held at Vassar College, N.Y., in November, 1987; these "Poughkeepsie Principles" were further elaborated in a series of design documents. The Guidelines, say document TEI ED P1, should:

- suffice to represent the textual features needed for research
- be simple, clear, and concrete
- be easy for researchers to use without special-purpose software
- allow the rigorous definition and efficient processing of texts
- provide for user-defined extensions
- conform to existing and emergent standards

The world of scholarship is large and divers. For the Guidelines to have wide acceptability, it was important to ensure that:

- the common core of textual features be easily shared
- additional specialist features be easy to add to (or remove from) a text
- multiple parallel encodings of the same feature should be possible
- the richness of markup should be user-defined, with a very small minimal requirement
- adequate documentation of the text and its encoding should be provided.

This tutorial gives only a flavour of what is possible, using only the basic core of the TEI Scheme. For full information, the published Guidelines should be consulted.

2 A Short Example

In this example, we first demonstrate how a passage of prose might be typed into a computer by someone aware of the need to be faithful to typographic appearances, but with little sense of the purpose of mark-up. In an ideal world, such output might be generated by a very accurate optical scanner. It attempts to be faithful to the appearance of the printed text, by retaining the original line breaks, by introducing blanks to represent the layout of the original headings and page breaks, and so forth. Where characters not available on the keyboard are needed (such as the accented letter *a* in *faál* or the long dash), it attempts to mimic their appearance. Such tricks are rarely portable and for analytic purposes, their use introduces needless complications.

```
                         CHAPTER 38

READER, I married him. A quiet wedding we had: he and I, the par-
son and clerk, were alone present. When we got back from church, I
went into the kitchen of the manor-house, where Mary was cooking
the dinner, and John cleaning the knives, and I said -
  'Mary, I have been married to Mr Rochester this morning.' The
housekeeper and her husband were of that decent, phlegmatic
```

```
order of people, to whom one may at any time safely communicate a
remarkable piece of news without incurring the danger of having
one's ears pierced by some shrill ejaculation and subsequently stunned
by a torrent of wordy wonderment. Mary did look up, and she did
stare at me; the ladle with which she was basting a pair of chickens
roasting at the fire, did for some three minutes hang suspended in air,
and for the same space of time John's knives also had rest from the
polishing process; but Mary, bending again over the roast, said only -
   'Have you, miss? Well, for sure!'
   A short time after she pursued, 'I seed you go out with the master,
but I didn't know you were gone to church to be wed'; and she
basted away. John, when I turned to him, was grinning from ear to
ear.
   'I telled Mary how it would be,' he said: 'I knew what Mr Ed-
ward' (John was an old servant, and had known his master when he
was the cadet of the house, therefore he often gave him his Christian
name) - 'I knew what Mr Edward would do; and I was certain he
would not wait long either: and he's done right, for aught I know. I
wish you joy, miss!' and he politely pulled his forelock.
   'Thank you, John. Mr Rochester told me to give you and Mary
this.'
   I put into his hand a five-pound note.  Without waiting to hear
more, I left the kitchen. In passing the door of that sanctum some time
after, I caught the words -
   'She'll happen do better for him nor ony o' t' grand ladies.' And
again, 'If she ben't one o' th' handsomest, she's noan faa\l, and varry
good-natured; and i' his een she's fair beautiful, onybody may see
that.'
   I wrote to Moor House and to Cambridge immediately, to say what
I had done: fully explaining also why I had thus acted. Diana and

                            474

                 JANE EYRE                      475

Mary approved the step unreservedly. Diana announced that she
would just give me time to get over the honeymoon, and then she
would come and see me.
   'She had better not wait till then, Jane,' said Mr Rochester, when I
read her letter to him; 'if she does, she will be too late, for our honey-
moon will shine our life long: its beams will only fade over your
grave or mine.'
   How St John received the news I don't know: he never answered
the letter in which I communicated it: yet six months after he wrote
to me, without, however, mentioning Mr Rochester's name or allud-
ing to my marriage. His letter was then calm, and though very serious,
kind. He has maintained a regular, though not very frequent correspond-
ence ever since: he hopes I am happy, and trusts I am not of those who
live without God in the world, and only mind earthly things.
```

This transcription suffers from a number of shortcomings:

- the page numbers and running titles are intermingled with the text in a way which makes it difficult for software to disentangle them
- no distinction is made between single quotation marks and apostrophe, so it is difficult to know exactly which passages are in direct speech
- the preservation of the copy text's hyphenation means that simple-minded search programs will not find the broken words
- the accented letter in *faál* and the long dash have been rendered by ad hoc keying conventions which follow no standard pattern and will be processed correctly only if the transcriber remembers to mention them in the documentation

- paragraph divisions are marked only by the use of white space, and hard carriage returns have been introduced at the end of each line. Consequently, if the size of type used to print the text changes, reformatting will be problematic.

We now present the same passage, as it might be encoded using the TEI Guidelines. As we shall see, there are many ways in which this encoding could be extended, but as a minimum, the TEI approach allows us to represent the following distinctions:

- Paragraph divisions are now marked explicitly
- Apostrophes are distinguished from quotation marks
- Entity references are used for the accented letter and the long dash
- Page divisions have been marked with an empty <PB> element alone.
- To simplify searching and processing, the lineation of the original has not been retained and words broken by typographic accident at the end of a line have been re-assembled without comment.
- For convenience of proof reading, a new line has been introduced at the start of each paragraph, but the indentation is removed.

```
<pb n='474'>
<div1 type=chapter n='38'>
<p>Reader, I married him.  A quiet wedding we had: he and I, the parson
and clerk, were alone present.  When we got back from church, I went
into the kitchen of the manor-house, where Mary was cooking the dinner,
and John cleaning the knives, and I said &dash;
<p><q>Mary, I have been married to Mr Rochester this morning.</q>
The housekeeper and her husband were of that decent, phlegmatic
order of people, to whom one may at any time safely communicate
a remarkable piece of news without incurring the danger of
having one's ears pierced by some shrill ejaculation and
subsequently stunned by a torrent of wordy wonderment.  Mary did
look up, and she did stare at me; the ladle with which she was
basting a pair of chickens roasting at the fire, did for some
three minutes hang suspended in air, and for the same space of
time John's knives also had rest from the polishing process; but
Mary, bending again over the roast, said only &dash;
<p><q>Have you, miss? Well, for sure!</q>
<p>A short time after she pursued, <q>I seed you go out with the
master, but I didn't know you were gone to church to be wed</q>;
and she basted away.  John, when I turned to him, was grinning
from ear to ear.  <q>I telled Mary how it would be,</q> he said:
<q>I knew what Mr Edward</q> (John was an old servant, and had
known his master when he was the cadet of the house, therefore
he often gave him his Christian name) &dash; <q>I knew what Mr
Edward would do; and I was certain he would not wait long
either: and he's done right, for aught I know.  I wish you joy,
miss!</q> and he politely pulled his forelock.
<p><q>Thank you, John.  Mr Rochester told me to give you and Mary this.</q>
<p>I put into his hand a five-pound note.  Without waiting to hear
more, I left the kitchen.  In passing the door of that sanctum some
time after, I caught the words &dash;
<p><q>She'll happen do better for him nor ony o' t' grand ladies.</q> And
again, <q>If she ben't one o' th' handsomest,
she's noan fa&agrave;l, and varry good-natured; and i' his een
she's fair beautiful, onybody may see that.</q>
<p>I wrote to Moor House and to Cambridge immediately, to say what I
had done: fully explaining also why I had thus acted.  Diana and
<pb n='475'> Mary approved the step unreservedly.
Diana announced that she would just give me time to get over the
honeymoon, and then she would come and see me.
<p><q>She had better not wait till then, Jane,</q> said Mr
```

```
Rochester, when I read her letter to him; <q>if she does, she
will be too late, for our honeymoon will shine our life long:
its beams will only fade over your grave or mine.</q>
<p>How St John received the news I don't know: he never answered the
letter in which I communicated it: yet six months after he wrote to me,
without, however, mentioning Mr Rochester's name or alluding to my
marriage.  His letter was then calm, and though very serious, kind.  He
has maintained a regular, though not very frequent correspondence ever
since: he hopes I am happy, and trusts I am not of those who live
without God in the world, and only mind earthly things.
```

The decision to focus on Brontë's text, rather than on the printing of it in this particular edition, is one aspect of a fundamental encoding issue: that of selectivity. An encoding makes explicit only those textual features of importance to the encoder. It is not difficult to think of ways in which the encoding of even this short passage might readily be extended. For example:

- a regularised form of the passages in dialect could be provided
- footnotes glossing or commenting on any passage could be added
- pointers linking parts of this text to others could be added
- proper names of various kinds could be distinguished from the surrounding text
- detailed bibliographic information about the text's provenance and context could be prefixed to it

The TEI-recommended way of carrying all of these out is described in the remainder of this tutorial. The TEI scheme as a whole also provides for an enormous range of other possibilities, of which we cite only a few:

- representation of a linguistic analysis of the passage as sentences, clauses, words etc. with associated category codes
- segmentation of the text into narrative or discourse units
- representation of complex alignment or linkage patterns, for example between this text and some analytic scheme, or between the text and a one or more translations of it
- detailed analysis of the components of names
- detailed meta-information providing thesaurus-style information about the text's origins or topics
- external links, for example to associate passages with images or sound held on other media
- information about the printing history or manuscript variations exhibited by a particular series of versions of the text

For recommendations on these and many other possibilities, the full Guidelines should be consulted.

3 The Structure of a TEI Text

All TEI-conformant texts contain (a) a *TEI header* (marked up as a <TEIHEADER> element) and (b) the transcription of the text proper (marked up as a <TEXT> element).

The TEI header provides information analogous to that provided by the title page of a printed text. It has up to four parts: a bibliographic description of the machine-readable text, a description of the way it has been encoded, a non-bibliographic description of the text (a *text profile*), and a revision history. The header is described in more detail in section 14 (on page 28).

A TEI text may be *unitary* (a single work) or *composite* (a collection of single works, such as an anthology): in this document we discuss only unitary works. A text may have an optional *front* or *back*. In between is the *body* of the text, which, in the case of a composite text, may consist of *group*s, each containing more groups or texts.

A typical TEI text will be encoded using an overall structure like this:

```
<TEI.2>
    <teiHeader> [ TEI Header information ]  </teiHeader>
    <text>
        <front> [ front matter ... ]   </front>
        <body>  [ body of text ... ]   </body>
        <back>  [ back matter ...  ]   </back>
    </text>
</TEI.2>
```

A typical composite text might have a structure like the following:

```
<TEI.2>
    <teiHeader> [ header information for the composite ] </teiHeader>
    <text>
        <front> [ front matter for the composite  ]      </front>
        <group>
           <text>
              <front> [ front matter for the first text ] </front>
              <body>  [ body of first text ... ]          </body>
              <back>  [ back matter of first text... ]    </back>
           </text>
           <text>
              <front> [ front matter for the second text]  </front>
              <body>  [ body of second text ... ]          </body>
              <back>  [ back matter of second text... ]    </back>
           </text>
        </group>
        <back>      [ back matter for the composite  ]      </back>
    </text>
</TEI.2>
```

It is also possible to define a composite of TEI texts, each with its own header. Such a collection is known as a *TEI corpus*, and may itself have a header:

```
<teiCorpus>
    <teiHeader> [ header information for the corpus ]  </teiHeader>
    <TEI.2>
       <teiHeader> [ header information  for first text]  </teiHeader>
       <text>      [ first text in corpus ]               </text>
    </TEI.2>
    <TEI.2>
       <teiHeader> [ header information  for second text]  </teiHeader>
       <text>      [ second text in corpus ]               </text>
    </TEI.2>
</teiCorpus>
```

In the remainder of this tutorial, we discuss only simple text structures. The discussion in each case consists of a short list of relevant TEI *elements* with a brief definition of each, followed by definitions for any *attributes* specific to that element. In most cases, short examples are also given.

4 Encoding the Basic Structure

As indicated above, a simple TEI document at the textual level consists of the following elements:

<FRONT> contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found before the start of a text proper.

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

<BACK> contains any appendixes, etc. following the main part of a text.

Elements specific to front and back matter are described below in section 12 (on page 24).

4.1 Text-Division Elements

The body of a text may be just a series of paragraphs, or these paragraphs may be grouped together into into chapters, sections, subsections, etc. In the former case, each paragraph is tagged using the <P> tag. In the latter case, the <BODY> may be divided either into a series of <DIV1> elements, or into a series of <DIV> elements, either of which may be further subdivided, as discussed below:

<P> marks paragraphs in prose.

<DIV> contains a subdivision of the front, body, or back 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).

When structural subdivisions smaller than a <DIV1> are necessary, a <DIV1> may be divided into <DIV2> elements, a <DIV2> into smaller <DIV3> elements, etc., down to the level of <DIV7>. If more than seven levels of structural division are present, one must either modify the TEI tag set to accept <DIV8> etc., or else use the unnumbered <DIV> element: a <DIV> may be subdivided by smaller <DIV> elements, without limit to the depth of nesting.

All the elements for structural divisions take the following three attributes:

TYPE This indicates the conventional name for this category of text division. Its value will typically be “Book”, “Chapter”, “Poem”, etc. Other possible values include “Group” for groups of poems etc. treated as a single unit, “Sonnet”, “Speech”, and “Song”. Note that the TYPE attribute need be supplied only for the first <DIV1>, <DIV2>, etc., in a text as its value is assumed to apply for all subsequent <DIV1>s within the same <BODY>.

ID This specifies a unique identifier for the division, which may be used for cross references and for commentary. It is probably a good idea to provide an ID attribute for every structural unit in a text, and to derive the ID values in some systematic way, for example by appending a section number to a short code for the title of the work in question, as in the examples below.

N The N attribute specifies a mnemonic short name or number for the division, which can be used to identify it in preference to the ID. If a conventional form of reference or abbreviation for the parts of a work already exists (such as the book/chapter/verse pattern of Biblical citations), the N attribute is the place to record it.

The attributes ID and N , indeed, are so widely useful that they are allowed on any element in any TEI DTD: they are *global attributes*.

The value of every ID attribute must be unique within a document. One simple way of ensuring that this is so is to make it reflect the hierarchic structure of the document. For example, Smith's Wealth of Nations as first published consists of five books, each of which is divided into chapters, while some chapters are further subdivided into parts. We might define ID values for this structure as follows:

```
<div1 id=WN1 n='I' type='book'>
  <div2 id=WN101 n='I.1' type='chapter'>
   ... </div2>
  <div2 id=WN102 n='I.2' type='chapter'>
   ... </div2>
   ...
  <div2 id=WN110 n='I.10' type='chapter'>
     <div3 id=WN1101 n='I.10.1' type=part>
      ... </div3>
     <div3 id=WN1102 n='I.10.2' type=part>
      ... </div3>
  </div2>
  ...
</div1>
<div1 id=WN2 n='II' type='book'>
   ....
</div1>
...
```

A different numbering scheme may be used for ID and N attributes: this is often useful where a canonical reference scheme is used which does not tally with the structure of the work. For example, in a novel divided into books each containing chapters, where the chapters are numbered sequentially through the whole work, rather than within each book, one might use a scheme such as the following:

```
<div1 id=TS01 n='1' type='Volume'>
   <div2 id=TS011 n='1' type='Chapter'>
      ...
   <div2 id=TS012 n='2'>
      ...
</div1>
<div1 id=TS02 n='2' type='Volume'>
   <div2 id=TS021 n='3'type='Chapter'>
      ...
   <div2 id=TS022 n='4'>
      ...
</div1>
```

Here the work has two volumes, each containing two chapters. The chapters are numbered conventionally 1 to 4, but the id values specified allow them to be regarded additionally as if they were numbered 1.1, 1.2, 2.1, 2.2.

4.2 Headings and Closings

Every <DIV>, <DIV1>, <DIV2>, etc. may have a title or heading at its start, and (less commonly) a closing such as "End of Chapter 1". The following elements may be used to transcribe them:

<HEAD> contains any heading, for example, the title of a section, or the heading of a list or glossary.

<TRAILER> contains a closing title or footer appearing at the end of a division of a text.

Whether or not headings and trailers are included in a transcription is a matter for the individual transcriber to decide. Where a heading is completely regular (for example 'Chapter 1') or has been given as an attribute value (e.g. <DIV1 TYPE='CHAPTER' N=1>), it may be omitted; where it contains otherwise unrecoverable text it should always be included. For example, the start of Hardy's Under the Greenwood Tree might be encoded as follows:

```
<div1 id=UGT1 n='Winter' type='Part'>
<div2 id=UGT11 n='1' type='Chapter'>
<head>Mellstock-Lane</head>
<p>To dwellers in a wood almost every species of tree ...
```

4.3 Paragraphs

In prose, paragraphs should be tagged with the <P> tag:

<P> marks paragraphs in prose.

For example:

```
<body>
<p>I fully appreciate Gen. Pope's splendid achievements with their
invaluable results; but you must know that Major Generalships in the
Regular Army, are not as plenty as blackberries.
</p>
</body>
```

4.4 Verse and Drama

The following elements are provided for the encoding of structures specific to verse and drama:

<L> contains a single, possibly incomplete, line of verse. Attributes include:

PART specifies whether or not the line is metrically complete. Legal values are:

- F the final part of an incomplete line
- Y the line is metrically incomplete
- N either the line is complete, or no claim is made as to its completeness
- I the initial part of an incomplete line
- M a medial part of an incomplete line

<LG> contains a group of verse lines functioning as a formal unit e.g. a stanza, refrain, verse paragraph, etc.

<SP> An individual speech in a performance text, or a passage presented as such in a prose or verse text. Attributes include:

WHO identifies the speaker of the part by supplying an ID.

<SPEAKER> A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.

<STAGE> contains any kind of stage direction within a dramatic text or fragment. Attributes include:

TYPE indicates the kind of stage direction. Suggested values include:

- `entrance` describes an entrance.
- `exit` describes an exit.
- `modifier` gives some detail about a character.
- `location` describes a location.
- `delivery` describes how a character speaks.
- `setting` describes a setting.
- `mixed` more than one of the above
- `business` describes stage business.

`novelistic` is a narrative, motivating stage direction.

This example shows a stanza divided between two speakers:

```
<sp><speaker>First voice</>
<lg type=stanza part=I>
<l>But why drives on that ship so fast
<l>Withouten wave or wind?
</lg>
<sp><speaker>Second Voice</speaker>
<lg part=F>
<l>The air is cut away before.
<l>And closes from behind.
</lg>
```

In this example, each speech consists of a sequence of verse lines, some of them incomplete.

```
<div1 type ='Act' n='I'><head>ACT I</head>
<div2 type ='Scene' n='1'><head>SCENE I</head>
<stage rend=italic>
Enter Barnardo and Francisco, two Sentinels, at several doors</stage>
<sp><speaker>Barn<l part=Y>Who's there?
<sp><speaker>Fran<l>Nay, answer me. Stand and unfold yourself.
<sp><speaker>Barn<l part=i>Long live the King!
<sp><speaker>Fran<l part=m>Barnardo?
<sp><speaker>Barn<l part=f>He.
<sp><speaker>Fran<l>You come most carefully upon your hour.
```

This example shows how dialogue presented in a prose work as if it were drama should be encoded. It also demonstrates the use of the WHO attribute to bear a code identifying the speaker of the piece of dialogue concerned:

```
<sp><speaker who=OPI>The reverend Doctor Opimiam</speaker>
<p>I do not think I have named a single unpresentable fish.
<sp><speaker who=GRM>Mr Gryll</speaker>
<p>Bream, Doctor: there is not much to be said for bream.
<sp><speaker who=OPI>The Reverend Doctor Opimiam</speaker>
<p>On the contrary, sir, I think there is much to be said for him.
In the first place....
<p>Fish, Miss Gryll - I could discourse to you on fish by the hour:
but for the present I will forbear...
</sp>
```

5 Page and Line Numbers

Page breaks, column breaks, and line breaks may be marked with the following elements. All are empty elements.

<PB> marks the boundary between one page of a text and the next in a standard reference system. Attributes include:

ED indicates the edition or version in which the page break is located at this point

<LB> marks the start of a new (typographic) line in some edition or version of a text. Attributes include:

ED indicates the edition or version in which the line break is located at this point

<CB> marks the boundary between one column of a text and the next in a standard reference system. Attributes include:

ED indicates the edition or version in which the column break is located at this point

These elements mark a single point in the text, not a span of text. The global N attribute may be used to supply the number of the page, column, or line beginning at the tag.

When working from a paginated original, it is often useful to record its pagination, if only to simplify later proof-reading. Recording the line breaks may be useful for the same reason; treatment of end-of-line hyphenation in printed source texts will require some consideration.

If pagination, etc., are marked for more than one edition, specify the edition in question using the ED attribute, and supply as many tags are necessary. For example, in the following passage we indicate where the page breaks occur in two different editions ("ED1" and "ED2")

```
<p>I wrote to Moor House and to Cambridge immediately, to say what I
had done: fully explaining also why I had thus acted.  Diana and
<pb ed=ED1 n='475'> Mary approved the step unreservedly.
Diana announced that she would <pb ed=ED2 n='485'>just give me
time to get over the  honeymoon, and then she would come and see me.
```

6 Marking Highlighted Phrases

6.1 Changes of Typeface, etc.

Highlighted words or phrases are those made visibly different from the rest of the text, typically by a change of type font, handwriting style, or ink color, intended to draw the reader's attention to them.

It is not always possible and may not be considered desirable to interpret the reasons for such changes of rendering in a text. In such cases, the element <HI> may be used to mark highlighted text without making any claim as to its status.

Alternatively, where the cause for the highlighting can be identified with confidence, a number of other, more specific, elements are available.

<HI> marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.

<EMPH> marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.

<FOREIGN> identifies a word or phrase as belonging to some language other than that of the surrounding text.

<MENTIONED> marks words or phrases mentioned, not used.

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

<TITLE> contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles. Attributes include:

LEVEL indicates whether this is the title of an article, book, journal, series, or unpublished material. Legal values are:

- m monographic title (book, collection, or other item published as a distinct item, including single volumes of multi-volume works)
- s series title
- j journal title
- u title of unpublished material (including theses and dissertations unless published by a commercial press)
- a analytic title (article, poem, or other item published as part of a larger item)

TYPE classifies the title according to some convenient typology. Sample values include:

`abbreviated` abbreviated form of title

`main` main title

`subordinate` subtitle, title of part

`parallel` alternate title, often in another language, by which the work is also known

The global REND attribute should be used wherever necessary to specify details of how the highlighting was realized. For example, a foreign word rendered in bold might be tagged <FOREIGN REND='BOLD'>, and one in italic <FOREIGN REND='ITALIC'>.

Some features (notably quotations and glosses) may be found in a text either marked by highlighting, or with quotation marks. In either case, the elements <Q> and <GLOSS> (as discussed in the following section) should be used. If the rendition is to be recorded, use the global REND attribute.

As an example of the elements defined here, consider the following sentence:

> On the one hand the Nibelungenlied is associated with the new rise of romance of twelfth-century France, the *romans d'antiquiteé* , the romances of Chreétien de Troyes, and the German adaptations of these works by Heinrich van Veldeke, Hartmann von Aue, and Wolfram von Eschenbach.

Interpreting the role of the highlighting, the sentence might look like this:

```
On the one hand the <title>Nibelungenlied</title> is associated
with the new rise of romance of twelfth-century France, the
<foreign>romans d'antiquit&eacute;</foreign>, the romances of
Chr&eacute;tien de Troyes, ...
```

Describing only the appearance of the original, it might look like this:

```
On the one hand the <hi rend=italic>Nibelungenlied</hi>
is associated with the new rise of romance of twelfth-century
France, the <hi rend=italic>romans
d'antiquit&eacute;</hi>, the romances of
Chr&eacute;tien de Troyes, ...
```

6.2 Quotations and Related Features

Like changes of typeface, quotation marks are conventionally used to denote several different features within a text, of which the most frequent is quotation. When possible, we recommend that the underlying feature be tagged, rather than the simple fact that quotation marks appear in the text, using the following elements:

<Q> contains a quotation or apparent quotation — a representation of speech or thought marked as being quoted from someone else (whether in fact quoted or not); in narrative, the words are usually those of of a character or speaker; in dictionaries, <Q> may be used to mark real or contrived examples of usage. Attributes include:

TYPE may be used to indicate whether the quoted matter is spoken or thought, or to characterize it more finely. Sample values include:

`spoken` representation of direct speech, usually marked by quotation marks.

`thought` representation of thought, e.g. internal monologue.

WHO identifies the speaker of a piece of direct speech.

<QUOTE> contains a phrase or passage attributed by the narrator or author to some agency external to the text.

<MENTIONED> marks words or phrases mentioned, not used.

<SOCALLED> contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.

Here is a simple example of a quotation:

```
Few dictionary makers are likely to forget
Dr. Johnson's description of the
lexicographer as <q>a harmless drudge.</q>
```

To record how a quotation was printed (for example, *in-line* or set off as a *display* or *block quotation*), the REND attribute should be used. This may also be used to indicate the kind of quotation marks used.

Direct speech interrupted by a narrator can be represented simply by ending the quotation and beginning it again after the interruption, as in the following example:

```
<p><q>Who-e debel you?</q> &mdash; he at last said &mdash; <q>you
no speak-e, damme, I kill-e.</q>  And so saying, the lighted
tomahawk began flourishing about me in the dark.
```

Quotations may be accompanied by a reference to the source or speaker, using the WHO attribute, whether or not the source is given in the text, as in the following example:

```
<q who=Wilson>Spaulding, he came down into the office just this
day eight weeks with this very paper in his hand, and he
says:&mdash;<q who=Spaulding>I wish to the Lord, Mr. Wilson, that
I was a red-headed man.</q></q>
```

This example also demonstrates how quotations may be embedded within other quotations: the speaker (Wilson) quotes another speaker (Spaulding).

The creator of the electronic text must decide whether quotation marks are replaced by the tags or whether the tags are added and the quotation marks kept. If the quotation marks are removed from the text, the REND attribute may be used to record the way in which they were rendered in the copy text.

As with highlighting, it is not always possible and may not be considered desirable to interpret the function of quotation marks in a text in this way. In such cases, the tag <HI REND=QUOTED> might be used to mark quoted text without making any claim as to its status.

6.3 Foreign Words or Expressions

Words or phrases which are not in the main language of the text may be tagged as such in one of two ways. If the word or phrase is already tagged for some reason, the element indicated should bear a value for the global LANG attribute indicating the language used. Where there is no applicable element, the element <FOREIGN> may be used, again using the LANG attribute. For example:

```
John has real <foreign lang=fra>savoir-faire</foreign>.

Have you read <title lang=deu>Die Dreigroschenoper</title>?

<mentioned lang=fra>Savoir-faire</mentioned> is French for knowhow.

The court issued a writ of <term lang=lat>mandamus</term>.
```

As the last three examples show, the <FOREIGN> element should not be used to tag foreign words if some other more specific element such as <TITLE>, <MENTIONED>, or <TERM> applies.

7 Notes

All notes, whether printed as footnotes, endnotes, marginalia, or elsewhere, should be marked using the same element:

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

- TYPE describes the type of note.
- RESP 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.
  - `(initials)` note added by the individual indicated by the initials.
  - `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.
- PLACE indicates where the note appears in the source text. Sample values include:
  - `inline` note appears as a marked paragraph in the body of the text.
  - `left` note appears in left margin.
  - `foot` note appears at foot of page.
  - `right` note appears in right margin.
  - `app[aratus]` note appears in the apparatus at the foot of the page.
  - `interlinear` note appears between lines of the text.
  - `end` note appears at end of chapter or volume.
- TARGET indicates the point of attachment of a note, or the beginning of the span to which the note is attached.
- 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.

Where possible, the body of a note should be inserted in the text, at the point at which its identifier or mark first appears. This may not be possible for example with marginalia, which may not be anchored to an exact location. For simplicity, it may be adequate to position marginal notes before the relevant paragraph or other element.

The N attribute may be used to supply the number or identifier of a note if this is required.

Examples:

```
Collections are ensembles of distinct
entities or objects of any sort.
<note place=foot n=1>
We explain below why we use the uncommon term
<mentioned>collection</mentioned>
instead of the expected
<mentioned>set</mentioned>.
Our usage corresponds to the <mentioned>aggregate</mentioned>
of many mathematical writings and to the sense of
<mentioned>class</mentioned> found
in older logical writings.
</note>
The elements ...

<stanza id=RAM609>
<note place=margin>The curse is finally expiated</note>
<l>And now this spell was snapt: once more
<l>I viewed the ocean green,
<l>And looked far forth, yet little saw
<l>Of what had else been seen &dash;
```

8 Cross References

Cross references within a text are marked using the following elements:

<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 as one or more SGML identifiers

<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 as one or more SGML identifiers

As well as TARGET , these elements share the following attributes:

TYPE categorizes the pointer in some respect, using any convenient set of categories.

RESP specifies the creator of the pointer.

TARGTYPE specifies the kinds of elements to which this pointer may point.

The <REF> element contains the text of the cross-reference itself; the <PTR> element is used for cross references signaled by symbols, icons, or other non-verbal means (or in document production systems, where the system can generate the correct verbal form of the cross reference).

The following two forms, for example, are logically equivalent (assuming we have documented somewhere the exact verbal form of cross references represented by <PTR> elements):

```
See especially <ref target=SEC12>section 12 on page 34</ref>.
```

```
See especially <ptr target=SEC12>.
```

The value of the TARGET attribute must be an SGML identifier in the current SGML document. This implies that the passage or phrase being pointed at must bear an identifier, and must therefore be tagged as an element of some kind. In the following example, the cross reference is to a <DIV1> element:

```
...
see especially <ptr target=SEC12>.
...
<div1 id=SEC12><head>Concerning Identifiers...
...
```

Because the ID attribute is global, any element in a document may be pointed to in this way. In the following example, a paragraph has been given an identifier so that it may be pointed at:

```
...
this is discussed in <ref target=pspec>the paragraph on links</ref>
...
<p id=pspec>Links may be made to any kind of element
...
```

These elements may also be used for hyper-textual links as well as for overt cross-references; The Guidelines also provide a related pair of elements which can point to locations outside the current SGML document, as well as a range of facilities for linking to elements not explicitly tagged in the current document.

9 Text Features Not Signalled Typographically

TEI P3 proposes elements for a number of features which have been found to be useful in a variety of analytic contexts, even though they are not generally signalled by any consistent or widely followed typographic convention. These include editorial interventions of a simple kind, dates, names and numbers.

9.1 Editorial Interventions

The following five elements are provided for simple editorial interventions. For full textual critical analysis, more detailed elements are described in the Guidelines (but not in this tutorial).

<SIC> contains text reproduced although apparently incorrect or inaccurate. Attributes include:

- CORR gives a correction for the apparent error in the copy text.
- RESP signifies the editor or transcriber responsible for suggesting the correction held as the value of the CORR attribute.
- CERT signifies the degree of certainty ascribed to the correction held as the value of the CORR attribute.

<CORR> contains the correct form of a passage apparently erroneous in the copy text. Attributes include:

- SIC gives the original form of the apparent error in the copy text.
- RESP signifies the editor or transcriber responsible for suggesting the correction held as the content of the <CORR> element.
- CERT signifies the degree of certainty ascribed to the correction held as the content of the <CORR> element.

<REG> contains a reading which has been regularized or normalized in some sense. Attributes include:

- ORIG gives the unregularized form of the text as found in the source copy.
- RESP identifies the individual responsible for the regularization of the word or phrase.

<GAP> indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible or inaudible. Attributes include:

- DESC gives a description of the omitted text.
- RESP indicates the editor, transcriber or encoder responsible for the decision not to provide any transcription of the text and hence the application of the <GAP> tag.

For example, the reading

```
 ... for his nose was as sharp as a pen and a' table of green feelds
```

might be rendered as

```
 ... for his nose was as sharp as a pen and <reg sic="a'">he</reg>
 <corr sic='table' ed=Gifford cert=0>babbl'd</corr> of green
<reg sic='feelds'>fields</reg>
```

9.2 Names, Dates, Numbers and Abbreviations

The TEI scheme provides tags for a large number of "data-like" features which may appear almost anywhere within almost any kind of text. These features may be of particular interest in a range of disciplines; they all relate to objects external to the text itself, such as the names of persons and places, numbers and dates. They also pose particular problems for many natural language processing (NLP) applications because of the variety of ways in which they may be presented within a text. The tags described here, by making such features explicit, reduce the complexity of processing texts containing them.

9.2.1 Names and referring strings

A *referring string* is a phrase which refers to some person, place, object etc. Two elements are provided to mark such strings:

<RS> contains a general purpose name or referring string. Attributes include:

> TYPE indicates more specifically the object referred to by the referencing string. Values might include "person", "place", "ship", "element" etc.

<NAME> contains a proper noun or noun phrase. Attributes include:

> TYPE indicates the type of the object which is being named by the phrase.

The TYPE attribute is used to distinguish amongst (for example) names of persons, places and organizations, where this is possible:

```
<q>My dear <rs type=person>Mr. Bennet</rs>, </q>
said his lady to him one day, <q>have you heard
that <rs type=place>Netherfield Park</rs> is let
at last?</q>
```

```
It being one of the principles of the
<rs type=organization>Circumlocution Office</rs> never,
on any account whatsoever, to give a straightforward answer,
<rs type=person>Mr Barnacle</rs> said, <q>Possibly.</q>
```

As the following example shows, the <RS> element may be used for any reference to a person, place, etc, not necessarily one in the form of a proper noun or noun phrase.

```
<q>My dear <rs type=person>Mr. Bennet</rs>,</q>
said <rs type=person>his lady</rs> to him
one day...
```

The <NAME> element by contrast is provided for the special case of referencing strings which consist only of proper nouns; it may be used synonymously with the <RS> element, or nested within it if a referring string contains a mixture of common and proper nouns.

Simply tagging something as a name is generally not enough to enable automatic processing of personal names into the canonical forms usually required for reference purposes. The name as it appears in the text may be inconsistently spelled, partial, or vague. Moreover, name prefixes such as *van* or *de la* , may or may not be included as part of the reference form of a name, depending on the language and country of origin of the bearer.

The following attributes are also available for these and similar elements to help overcome these difficulties:

KEY provides an alternative identifier for the object being named, such as a database record key.

REG gives a normalized or regularized form of the name used.

The KEY attribute may be useful as a means of gathering together all references to the same individual or location scattered throughout a document:

```
<q>My dear <rs type=person key=BENM1>Mr. Bennet</rs>,
</q> said <rs type=person key=BENM2>his lady</rs>
to him one day, <q>have you heard that
<rs type=place key=NETP1>Netherfield Park</rs>
is let at last?</q>
```

This use should be distinguished from the case of the REG (regularization) attribute, which provides a means of marking the standard form of a referencing string as demonstrated below:

```
<name type=person key=WADLM1 reg='de la Mare, Walter'>
   Walter de la Mare
</name>
was born at
<name key=Ch1 type=place>Charlton</name>, in
<name key=KT1 type=county>Kent</name>, in 1873.
```

More detailed tagging of the components of proper names is also possible, using the additional tag set for names and dates.

9.2.2 Dates and times

Tags for the more detailed encoding of times and dates include the following:

<DATE> contains a date in any format. Attributes include:

- CALENDAR indicates the system or calendar to which the date belongs.
- VALUE gives the value of the date in some standard form, usually yyyy-mm-dd.

<TIME> contains a phrase defining a time of day in any format. Attributes include:

- VALUE gives the value of the time in a standard form.

<DATERANGE> contains two dates or another phrase delimiting a time period. Attributes include:

- CALENDAR indicates the system or calendar to which the date belongs.
- FROM indicates the starting point of the period in standard form.
- TO indicates the ending point of the period in standard form.
- EXACT indicates the precision to be attached to either or both dates specified. Sample values include:
  - `both` both dates are exact
  - `to` the TO date is exact
  - `none` both dates are approximate or unspecified
  - `from` the FROM date is exact

The VALUE attribute specifies a normalized form for the date or time, using a recognized format such as ISO 8601. Partial dates or times (e.g. “1990,”“September 1990”, “twelvish”) can usually be expressed by simply omitting a part of the value supplied; alternatively imprecise dates or times (for example “early August”, “some time after ten and before twelve”) may be expressed as date or time ranges. If either end of the date or time range is known to be accurate, (for example, “at some time before 1230”, “a few days after Hallowe’en”) the EXACT attribute may be used to specify this.

Examples:

```
<date value='1980-02-21'>21 Feb 1980</date>
<date value='1990'>1990</date>
<date value='1990-09'>September 1990</date>

Given on the <date value='1977-06-12'>Twelfth Day of June
in the Year of Our Lord One Thousand Nine Hundred and
Seventy-seven of the Republic the Two Hundredth and first
and of the University the Eighty-Sixth.</date>

Those five years &mdash; <dateRange from=1918 to=1923>
1918 to 1923</dateRange> &mdash had been, he suspected,
somehow very important.

<l>specially when it's nine below zero
<l>and <time value='15:00'>three o'clock in the afternoon</time>
```

9.2.3 Numbers

Numbers can be written with either letters or digits ("twenty-one", "xxi", and "21") and their presentation is language-dependent (e.g. English "5th" becomes Greek "5."; English "123,456.78" equals French "123.456,78"). In NLP or MT applications, it is often helpful to distinguish them from other, more "lexical" parts of the text. In other applications, the ability to standardize a number's value is important. The <NUM> element provides this possibility:

<NUM> contains a number, written in any form. Attributes include:

- TYPE indicates the type of numeric value. Suggested values include:
  - `fraction` fraction, e.g. "one half" or "three halves"
  - `ordinal` ordinal number, e.g. "21st"
  - `percentage` e.g. "ten percent"
  - `cardinal` absolute number, e.g. "21", "21.5"
- VALUE supplies the value of the number in an application-dependent standard form.

For example:

```
<num value='33'>xxxiii</num>
<num type=cardinal value='21'>twenty-one</num>
<num type=percentage value='10'>ten percent</num>
<num type=percentage value='10'>10%</num>
<num type=ordinal value='5'>5th</num>
```

9.2.4 Abbreviations and their Expansion

Like names, dates, and numbers, abbreviations may be transcribed as they stand or expanded; they may be left unmarked, or encoded using the following element:

<ABBR> contains an abbreviation of any sort. Attributes include:

- EXPAN gives an expansion of the abbreviation.
- TYPE allows the encoder to classify the abbreviation according to some convenient typology. Sample values include:
  - `contraction` the abbreviation omits some letter(s) in the middle.
  - `title` the abbreviation is for a title of address (Dr, Ms, Mr, ...)
  - `geographic` the abbreviation is for a geographic name.
  - `suspension` the abbreviation provides the first letter(s) of the word or phrase, omitting the remainder.
  - `organization` the abbreviation is for the name of an organization.
  - `brevigraph` the abbreviation comprises a special symbol or mark.

`superscription` the abbreviation includes writing above the line.

`acronym` the abbreviation comprises the initial letters of the words of a phrase.

The <ABBR> element is useful as a means of distinguishing semi-lexical items such as acronyms or jargon:

```
We can sum up the above discussion as follows:  the identity of a
<abbr>CC</abbr> is defined by that calibration of values which
motivates the elements of its <abbr>GSP</abbr>;

Every manufacturer of <abbr>3GL</abbr> or <abbr>4GL</abbr>
languages is currently nailing on <abbr>OOP</abbr> extensions
```

The TYPE attribute may be used to distinguish types of abbreviation by their function, and the EXPAN attribute may be used to supply an expansion:

```
<name><abbr type=title expan='Doctor'>Dr.</abbr>
<abbr type=initial expan='Marilyn'>M.</abbr>
Deegan</name>
is the Director of the
<abbr expan='Computers in Teaching Initiative' type=acronym>
CTI</abbr> Centre for Textual Studies.
```

This element is also particularly useful where manuscript materials in which abbreviation is very frequent are being transcribed.

10 Lists

The element <LIST> is used to mark any kind of *list*. A list is a sequence of text items, which may be ordered, unordered, or a glossary list. Each item may be preceded by an item label (in a glossary list, this label is the term being defined):

<LIST> contains any sequence of items organized as a list. Attributes include:

TYPE describes the form of the list. Suggested values include:

`ordered` list items are numbered or lettered.

`bulleted` list items are marked with a “bullet” or other typographic device.

`gloss` each list item glosses some term or concept, which is given by a <LABEL> element preceding the list item.

`simple` list items are not numbered or bulleted.

<ITEM> contains one component of a list.

<LABEL> contains the label associated with an item in a list; in glossaries, marks the term being defined.

Individual list items are tagged with <ITEM>. The first <ITEM> may optionally be preceded by a <HEAD>, which gives a heading for the list. The numbering of a list may be omitted (if reconstructible), indicated using the N attribute on each item, or (rarely) tagged as content using the <LABEL> element. The following are all thus equivalent: Kittred

```
<list>
<head>A short list</head>
<item>First item in list.</item>
<item>Second item in list.</item>
<item>Third item in list.</item>
</list>

<list>
```

```
<head>A short list</head>
<item n=1>First item in list.</item>
<item n=2>Second item in list.</item>
<item n=3>Third item in list.</item>
</list>

<list>
<head>A short list</head>
<label>1</label><item>First item in list.</item>
<label>2</label><item>Second item in list.</item>
<label>3</label><item>Third item in list.</item>
</list>
```

The styles should not be mixed in the same list.

A simple two-column table may be treated as a *glossary list*, tagged <LIST TYPE=GLOSS>. Here, each item comprises a *term* and a *gloss*, marked with <LABEL> and <ITEM> respectively. These correspond to the elements <TERM> and <GLOSS>, which can occur anywhere in prose text. The two columns may be headed with <HEADLABEL> and <HEADITEM> elements, as below:

```
<list type=gloss>
<head>Vocabulary</head>
<headLabel>Middle English</headLabel>
<headItem>New English</headItem>
<label lang=enm>nu        <item>now
<label lang=enm>lhude     <item>loudly
<label lang=enm>bloweth   <item>blooms
<label lang=enm>med       <item>meadow
<label lang=enm>wude      <item>wood
<label lang=enm>awe       <item>ewe
<label lang=enm>lhouth    <item>lows
<label lang=enm>sterteth  <item>bounds, frisks
<label lang=enm>verteth   <item lang=lat>pedit
<label lang=enm>murie     <item>merrily
<label lang=enm>swik      <item>cease
<label lang=enm>naver     <item>never
</list>
```

Where the internal structure of a list item is more complex, it may be preferable to regard the list as a *table*, for which special purpose tagging is defined in an additional TEI tag set.

Lists of whatever kind can, of course, nest within list items to any depth required. Here, for example, a glossary list contains two items, each of which is itself a simple list:

```
<list type=gloss><label>EVIL</label>
<item><list type=simple>
   <item>I am cast upon a horrible desolate island, void
          of all hope of recovery.</item>
   <item>I am singled out and separated as it were from
         all the world to be miserable.</item>
   <item>I am divided from mankind &mdash a solitaire; one
           banished from human society.</item>
     </list> <!- end of first nested list -></item>
<label>GOOD</label>
<item><list type=simple>
     <item>But I am alive; and not drowned, as all my
               ship's company were.</item>
     <item>But I am singled out, too, from all the ship's
              crew, to be spared from death...</item>
     <item>But I am not starved, and perishing on a barren place,
             affording no sustenances....</item>
     </list><!- end of second nested list -></item>
</list><!- end of glossary list ->
```

A list need not necessarily be displayed in list format. For example,

```
On those remote pages it is written that animals are
divided into <list><item n=’a’>those that belong to the
Emperor,<item n=’b’> embalmed ones, <item n=’c’> those
that are trained, <item n=’d’> suckling pigs, <item n=’e’>
mermaids,  <item n=’f’> fabulous ones, <item n=’g’> stray
dogs, <item n=’h’> those that are included in this
classification, <item n=’i’> those that tremble as if they
were mad, <item n=’j’> innumerable ones, <item n=’k’> those
drawn with a very fine camel’s-hair brush, <item n=’l’>
others, <item n=’m’> those that have just broken a flower
vase, <item n=’n’> those that resemble flies from a
distance.</list>
```

11 Bibliographic Citations

Bibliographic citations may be wholly absent from many works transcribed for research; where they occur, however, it is very useful to mark them for special handling. No full treatment of bibliographic references can or need be given here; the need is to mark the reference provided by the text, not to correct it.

The following elements are provided for use in transcribing bibliographic references; in most cases, only the titles of articles, books, and journals need be marked (since they will require special handling—font shifts—to print correctly). The other elements are provided for cases where particular interest attaches to such details, and for use in structured citations (on which see below). Their use is strictly optional.

<BIBL> contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

<AUTHOR> in a bibliographic reference, contains the name of the author(s), personal or corporate, of a work; the primary *statement of responsibility* for any bibliographic item.

<EDITOR> secondary *statement of responsibility* for a bibliographic item, for example the name of an individual, institution or organization, (or of several such) acting as editor, compiler, translator, etc.

<TITLE> contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles.

<SERIES> contains information about the series in which a book or other bibliographic item has appeared.

<IMPRINT> groups information relating to the publication or distribution of a bibliographic item.

<BIBLSCOPE> defines the scope of a bibliographic reference, for example as a list of pagenumbers, or a named subdivision of a larger work.

<PUBPLACE> contains the name of the place where a bibliographic item was published.

<PUBLISHER> provides the name of the organization responsible for the publication or distribution of a bibliographic item.

<DATE> contains a date in any format.

For example, the following editorial note might be transcribed as shown:

> He was a member of Parliament for Warwickshire in 1445, and died March 14, 1470 (according to Kittredge, Harvard Studies 5. 88ff).

```
He was a member of Parliament for Warwickshire in 1445, and died
March 14, 1470 (according to <bibl><author>Kittredge</author>,
<title>Harvard Studies</title> <biblScop>5. 88ff</biblScope></bibl>).
```

For the systematic encoding of bibliographies, P3 provides another element, <BIBLSTRUCT>, which requires its contents to be a bibliographic citation well formed according to the rules of the International Standard for Bibliographic Description (ISBD): the author, title, editor, series, imprint, and other information must all be tagged and provided in the order prescribed by ISBD.

For lists of bibliographic citations, the <LISTBIBL> element should be used; it may contain a series of <BIBL> or <BIBLSTRUCT> elements.

12 Front and Back Matter

12.1 Front Matter

For many purposes, particularly in older texts, the preliminary material such as title pages, prefatory epistles etc. may provide very useful additional linguistic or social information. P3 provides a set of recommendations for distinguishing the textual elements most commonly encountered in front matter, which are summarized here.

12.1.1 Title Page

The start of a title page should be marked with the element <TITLEPAGE>. All text contained on the page should be transcribed and tagged with the appropriate element from the following list:

<TITLEPAGE> contains the title page of a text, appearing within the front or back matter.

<DOCTITLE> contains the title of a document, including all its constituents, as given on a title page.

<TITLEPART> contains a subsection or division of the title of a work, as indicated on a title page. Attributes include:

- TYPE specifies the role of this subdivision of the title. Suggested values include:
  - `desc` descriptive paraphrase of the work included in title
  - `alt` alternative title of the work
  - `main` main title of the work
  - `sub` subtitle of the work

<EPIGRAPH> contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.

<DOCAUTHOR> contains the name of the author of the document, as given on the title page (often but not always contained in a <BYLINE>).

<BYLINE> contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.

<DOCEDITION> contains an edition statement as presented on a title page of a document.

<DOCIMPRINT> contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.

Typeface distinctions should be marked with the REND attribute when necessary, as described above. Very detailed description of the letter spacing and sizing used in ornamental titles is not as yet provided for by the Guidelines. Changes of language should be marked by appropriate use of the LANG attribute or the <FOREIGN> element, as necessary. Names, wherever they appear, should be tagged using the <NAME>, as elsewhere.

Two example title pages follow:

```
<titlePage rend=Roman>
  <docTitle>
    PARADISE REGAIN'D. A POEM In IV <hi>BOOKS</hi>.
    To which is added <title>SAMSON AGONISTES</title>.
  </docTitle>
  <byLine>The Author <docAuthor>JOHN MILTON</docAuthor></byline>
  <docImprint><name>LONDON</name>,
    Printed by <name>J.M.</name>
    for <name>John Starkey</name>
    at the <name>Mitre</name>
    in <name>Fleetstreet</name>,
    near <name>Temple-Bar.</name>
  </docImprint>
  <docDate>MDCLXXI</docDate>
</titlePage>

<titlePage>
  <docTitle>Lives of the Queens of England, from the Norman
    Conquest; with anecdotes of their courts.</docTitle>
  <titlePart>Now first published from Official Records
    and other authentic documents private as well as
    public.</titlePart>
  <docEdition>New edition, with corrections and
    additions</docEdition>
  <byline>By <docAuthor>Agnes Strickland</docAuthor></byline>
  <epigraph>
    <q>The treasures of antiquity laid up in old
       historic rolls, I opened.</q>
    <bibl>BEAUMONT</bibl>
  </epigraph>
  <docImprint>Philadelphia: Blanchard and Lea</docImprint>
  <docDate>1860.</docDate>
</titlePage>
```

12.1.2 Prefatory Matter

Major blocks of text within the front matter should be marked as <DIV> or <DIV1> elements; the following suggested values for the TYPE attribute may be used to distinguish various common types of prefatory matter:

`foreword` a text addressed to the reader, by the author, editor or publisher, possibly in the form of a letter.

`preface` a text addressed to the reader, by the author, editor or publisher, possibly in the form of a letter.

`dedication` a text (often a letter) addressed to someone other than the reader in which the author typically commends the work in hand to the attention of the person concerned.

`abstract` a prose argument summarizing the content of the work

`ack` Acknowledgements

`contents` a table of contents (typically this should be tagged as a <LIST>)

`frontispiece` a pictorial frontispiece, possibly including some text

Like any text division, those in front matter may contain low level structural or non-structural elements as described elsewhere. They will generally begin with a heading or title of some kind which should be tagged using the <HEAD> element. Epistles will contain the following additional elements:

<SALUTE> contains a salutation or greeting prefixed to a foreword, dedicatory epistle or other division of a text, or the salutation in the closing of a letter, preface, etc.

<SIGNED> contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text.

<BYLINE> contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.

<DATELINE> contains a brief description of the place, date, time, etc. of production of a letter, newspaper story, or other work, prefixed or suffixed to it as a kind of heading or trailer.

<OPENER> groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter.

<CLOSER> groups together dateline, byline, salutation, and similar phrases appearing as a final group at the end of a division, especially of a letter.

Epistles which appear elsewhere in a text will, of course, contain these same elements.

As an example, the dedication at the start of Milton's Comus should be marked up as follows:

```
<div type='dedication'>
<head>To the Right Honourable <name>JOHN Lord Viscount
BRACLY</name>, Son and Heir apparent to the Earl of
Bridgewater, &amp;c.</head>
<salute>MY LORD,</salute>

<p>THis <hi>Poem</hi>, which receiv'd its first occasion of
Birth from your Self, and others of your Noble Family ....
and as in this representation your attendant
<name>Thyrsis</name>, so now in all reall expression
<closer>
<salute>Your faithfull, and most humble servant</salute>
<signed><name>H. LAWES.</name></signed>
</closer>
</div>
```

12.2 Back Matter

Because of variations in publishing practice back matter can contain virtually any of the elements listed above for front matter, and the same elements should be used where this is so. Additionally, back matter may contain the following types of matter within the <BACK> element. Like the structural divisions of the body, these should be marked as <DIV> or <DIV1> elements, and distinguished by the following suggested values of the TYPE attribute:

`appendix` an appendix

`glossary` a list of words and definitions, typically in the form of a <LIST TYPE=GLOSS>

`notes` a series of <NOTE>s

`bibliography` a series of bibliographic references, typically in the form of a special bibliographic-list element <LISTBIBL>, whose items are individual <BIBL> or <BIBLSTRUCT> elements index a set of index entries, possibly represented as a structured list or glossary list, with optional leading <HEAD> and perhaps some paragraphs of introductory or closing text (TEI P3 defines other specialized elements for generating indices in document production)

colophon a description at the back of the book describing where, when, and by whom it was printed; in modern books it also often gives production details and identifies the type faces used

13 Character Sets, Diacritics, etc.

For those working with standard forms of the European languages, the TEI recommendations for character set use are simple. For local use, use whatever character set is supported by your machine and your software. If your software makes direct keyboard entry of special characters difficult, you may elect to define your own keyboarding conventions (for example to represent accented letters by typing the appropriate accent immediately after the letter, or by using special sequences unlikely to appear in normal text such as "aE" for "æ"). Global search and replace functions can then be used to turn these keyboard shorthands into the proper characters. If you work with non-Latin scripts and there is a standard transliteration scheme in your field (e.g. for ancient Greek the beta code of the Thesaurus Linguae Graecae), use it. Any transliteration used should be reversible (this rules out a surprising number of schemes commonly used in normal writing) and will be most usable if it requires no special ligatures, ties, or diacritics (this rules out a surprising number of the remainder).

For interchange of files among systems, use SGML entity references to replace all characters not in the following list of characters which almost always survive electronic interchange intact:

```
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
" % & ' ( ) * + , - . / : ; < = > ? _   (space)
```

Perhaps surprisingly, this list excludes the following characters which often do not survive transfer across national boundaries or over standard wide-area networks (if you're just going from your Mac to your PC, these characters will probably be safe):

```
 !  #  $    [    \  ]   ^   ` '  {   } |  ^
```

Less surprisingly, entity references must be used for all accented and extended-Latin characters, all non-Latin characters, and all symbols not on conventional computer keyboards.

You may use your own SGML entity names in TEI-conformant files, if you wish and if you provide standard SGML entity declarations for them, but the standard names (though long-winded) have the advantage of clarity; the characters intended are reasonably clear to any speaker of English who recognizes that a character is being named, often even without recourse to any list. This is not true of many other schemes for representing accented characters.

The entity names needed for the characters listed above as "unsafe" and for the accented characters of some major Western European languages are given below. Lists of public entity sets and their contents are available in any reference work on SGML: the following names are widely used, and are therefore recommended:

Unsafe characters: The following table shows recommended entity names for each of the "unsafe" characters listed above:

!	excl	#	num	$	dollar
[	lsqb	\	bsol	]	rsqb
^	circ	`	lsqo	'	grave
{	lcub	}	rcub	|	verbar
		^	tilde		

Digraphs Recommended entity names for digraphs are formed by appending the string "lig" to the letters forming the digraph. If a capitalized form is required, both letters are given in upper case (remember that case is usually significant in entity names). For example, the entity name for æ is "aelig", and that for Æis "AElig".

Diacritics and accents Recommended entity names for accented letters in most Western European languages are formed by appending one of the following strings to the letter bearing the accent, which may be in upper or lower case:

suffix	diacritic name	example
uml	umlaut or trema	`ä Ä` for ä Ä
acute	acute or stressed	`á Á` for á Á
grave	grave	`à À` for á Á
circ	circumflex	`â Â` for â Â
tilde	tilde	`ã Ã` for ã Ã

Consonants: The following table shows recommended entity names for some special consonants likely to be found in Western European languages:

c	ccedil	C	Ccedil
	eth		ETH
	thorn		THORN
ß	szlig		

Punctuation marks: The following table shows recommended entity names for some commonly found punctuation marks:

“	ldquo	”	rdquo
–	mdash	...	hellip

14 The Electronic Title Page

Every TEI text has a header which provides information analogous to that provided by the title page of printed text. The header is introduced by the element <teiHeader> and has four major parts:

<fileDesc> contains a full bibliographic description of an electronic file.

<encodingDesc> documents the relationship between an electronic text and the source or sources from which it was derived.

<profileDesc> provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.

<revisionDesc> summarizes the revision history for a file.

A corpus or collection of texts, which share many characteristics, may have one header for the corpus and individual headers for each component of the corpus. In this case the TYPE attribute indicates the type of header.

```
<teiHeader type=corpus>
```

introduces the header for corpus-level information.

Some of the header elements contain running prose which consists of one or more <P>s. Others are grouped:

- Elements whose names end in Stmt (for statement) usually enclose a group of elements recording some structured information.
- Elements whose names end in Decl (for declaration) enclose information about specific encoding practices.
- Elements whose names end in Desc (for description) contain a prose description.

14.1 The File Description

The <FILEDESC> element is mandatory. It contains a full bibliographic description of the file with the following elements:

<TITLESTMT> groups information about the title of a work and those responsible for its intellectual content.

<EDITIONSTMT> groups information relating to one edition of a text.

<EXTENT> describes the approximate size of the electronic text as stored on some carrier medium, specified in any convenient units.

<PUBLICATIONSTMT> groups information concerning the publication or distribution of an electronic or other text.

<SERIESSTMT> groups information about the *series*, if any, to which a publication belongs.

<NOTESSTMT> collects together any notes providing information about a text additional to that recorded in other parts of the bibliographic description.

<SOURCEDESC> supplies a bibliographic description of the copy text(s) from which an electronic text was derived or generated.

A minimal header has the following structure:

```
<teiHeader>
     <fileDesc>
          <titleStmt> ... </titleStmt>
          <publicationStmt> ... <publicationStmt>
          <sourceDesc> ... <sourceDesc>
     </fileDesc>
</teiHeader>
```

14.1.1 The title statement

The following elements can be used in the <TITLESTMT>:

<TITLE> contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles.

<AUTHOR> in a bibliographic reference, contains the name of the author(s), personal or corporate, of a work; the primary *statement of responsibility* for any bibliographic item.

<SPONSOR> specifies the name of a sponsoring organization or institution.

<FUNDER> specifies the name of an individual, institution, or organization responsible for the funding of a project or text.

<PRINCIPAL> supplies the name of the principal researcher responsible for the creation of an electronic text.

<RESPSTMT> supplies a statement of responsibility for someone responsible for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.

It is recommended that the title should distinguish the computer file from the source text, for example:

```
[title of source]: a machine readable transcription
[title of source]: electronic edition
A machine readable version of: [title of source]
```

The <RESPSTMT> element contains the following subcomponents:

<RESP> contains a phrase describing the nature of a person's intellectual responsibility.

<NAME> contains a proper noun or noun phrase.

Example:

```
<titleStmt>
     <title>Two stories by Edgar Allen Poe: a machine readable
               transcription</title>
     <author>Poe, Edgar Allen (1809-1849)
     <respStmt><resp>compiled by</resp>
     <name>James D. Benson</name></respStmt>
</titleStmt>
```

14.1.2 The edition statement

The <EDITIONSTMT> groups information relating to one edition of a text (where "edition" is used as elsewhere in bibliography), and may include the following elements:

<EDITION> describes the particularities of one edition of a text.

<RESPSTMT> supplies a statement of responsibility for someone responsible for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.

Example:

```
<editionStmt>
     <edition n=U2>Third draft, substantially revised
     <date>1987</date>
     </edition>
</editionStmt>
```

Determining exactly what constitutes a new edition of an electronic text is left to the encoder.

14.1.3 The extent statement

The <EXTENT> statement describe the approximate size of a file.
Example:

```
<extent>4532 bytes</extent>
```

14.1.4 The publication statement

The <PUBLICATIONSTMT> is mandatory. It may contain a simple prose description or groups of the elements described below:

<PUBLISHER> provides the name of the organization responsible for the publication or distribution of a bibliographic item.

<DISTRIBUTOR> supplies the name of a person or other agency responsible for the distribution of a text.

<AUTHORITY> supplies the name of a person or other agency responsible for making an electronic file available, other than a publisher or distributor.

At least one of these three elements must be present, unless the entire publication statement is in prose. The following elements may occur within them:

<PUBPLACE> contains the name of the place where a bibliographic item was published.

<ADDRESS> contains a postal or other address, for example of a publisher, an organization, or an individual.

<IDNO> supplies any standard or non-standard number used to identify a bibliographic item. Attributes include:

TYPE categorizes the number, for example as an ISBN or other standard series.

<AVAILABILITY> supplies information about the availability of a text, for example any restrictions on its use or distribution, its copyright status, etc. Attributes include:

STATUS supplies a code identifying the current availability of the text. Sample values include:

`restricted` the text is not freely available.

`unknown` the status of the text is unknown.

`free` the text is freely available.

<DATE> contains a date in any format.

Example:

```
<publicationStmt>
     <publisher>Oxford University Press</publisher>
     <pubPlace>Oxford</pubPlace> <date>1989</date>
     <idno type=ISBN> 0-19-254705-5</idno>
     <availability>Copyright 1989, Oxford University
          Press</availability>
</publicationStmt>
```

14.1.5 Series and Notes Statements

The <SERIESSTMT> groups information about the series, if any, to which a publication belongs. It may contain <TITLE>, <IDNO>, or <RESPSTMT> elements.

The <NOTESSTMT>, if used, contains one or more <NOTE> elements which contain a note or annotation. Some information found in the notes area in conventional bibliography has been assigned specific elements in the TEI scheme.

14.1.6 The Source Description

The <SOURCEDESC> is a mandatory element which records details of the source or sources from which the computer file is derived. it may contain simple prose or a bibliographic citation, using one or more of the following elements:

<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.

<LISTBIBL> contains a list of bibliographic citations of any kind.

For spoken texts, <SOURCEDESC> may also contain specialised <SCRIPTSTMT> and <RECORDINGSTMT> elements.

Examples:

```
<sourceDesc>
     <bibl>The first folio of Shakespeare, prepared by Charlton
          Hinman (The Norton Facsimile, 1968)</bibl>
</sourceDesc>

<sourceDesc>
     <scriptStmt id=CNN12>
     <bibl><author>CNN Network News
          <title>News headlines
          <date>12 Jun 1989
     </bibl>
     </scriptStmt>
</sourceDesc>
```

14.2 The Encoding Description

The <ENCODINGDESC> element specifies the methods and editorial principles which governed the transcription of the text. Its use is highly recommended. It may be prose description or may contain elements from the following list:

<PROJECTDESC> describes in detail the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.

<SAMPLINGDECL> contains a prose description of the rationale and methods used in sampling texts in the creation of a corpus or collection.

<EDITORIALDECL> provides details of editorial principles and practices applied during the encoding of a text.

<TAGSDECL> provides detailed information about the tagging applied to an SGML document.

<REFSDECL> specifies how canonical references are constructed for this text.

<CLASSDECL> contains one or more taxonomies defining any classificatory codes used elsewhere in the text.

14.2.1 Project and Sampling descriptions

Examples of <PROJECTDESC> and <SAMPLINGDESC>:

```
<encodingDesc>
     <projectDesc>Texts collected for use in the Claremont
          Shakespeare Clinic, June 1990.
     </projectDesc>
</encodingDesc>

<encodingDesc>
     <samplingDecl>Samples of 2000 words taken from the beginning
          of the text
     </samplingDecl>
</encodingDesc>
```

14.2.2 Editorial declarations

The <EDITORIALDECL> element may contain prose description or one or more of the following elements:

<CORRECTION> states how and under what circumstances corrections have been made in the text.

<NORMALIZATION> indicates the extent of normalization or regularization of the original source carried out in converting it to electronic form.

<QUOTATION> specifies editorial practice adopted with respect to quotation marks in the original.

<HYPHENATION> summarizes the way in which hyphenation in a source text has been treated in an encoded version of it.

<SEGMENTATION> describes the principles according to which the text has been segmented, for example into sentences, tone-units, graphemic strata, etc.

<INTERPRETATION> describes the scope of any analytic or interpretive information added to the text in addition to the transcription.

Example:

```
<editorialDecl>
     <interpretation>
          <p>The part of speech analysis applied throughout
               section 4 was added by hand and has not been
               checked
     <correction>
          <p>Errors in transcription controlled by using the
               WordPerfect spelling checker.
     <normalization source=W9>
          <p>All words converted to Modern American spelling
               using Webster’s 9th Collegiate dictionary.
     <quotation marks=all form=std>
          <p>All quotation marks converted to entity
               references &odq; and &cdq;
</editorialDecl>
```

14.2.3 Tagging, reference and classification declarations

The <TAGSDECL> element is used to provide detailed information about the SGML tags actually used within a text. It may contain a simple list of elements used, with a count for each, using the following special purpose elements:

<TAGUSAGE> supplies information about the usage of a specific element within a <TEXT>. Attributes include:

GI the name (generic identifier) of the element indicated by the tag.

OCCURS specifies the number of occurrences of this element within the text.

For example:

```
<tagsDecl>
 <tagUsage gi=text occurs=1>
 <tagUsage gi=p occurs=12>
 <tagUsage gi=hi occurs=6>
</tagsDecl>
```

This (imaginary) tags declaration would be appropriate for a text containing twelve paragraphs, within which six <HI> elements have been marked. Note that if the <TAGSDECL> element is used, it must contain a <TAGUSAGE> element for every element tagged in the associated text element.

The <REFSDECL> element is used to document the way in which any standard referencing scheme built into the encoding works. In its simplest form, it consists of prose description

Example:

```
<refsDecl>
     <p>The N attribute on each DIV1 and DIV2 contains the
     canonical reference for each such division in the form
     XX.yyy where XX is the book number in roman numeral and
     yyy is the section number in arabic
</refsDecl>
```

The <CLASSDECL> element groups together definitions or sources for any descriptive classification schemes used by other parts of the header. At least one such scheme must be provided, encoded using the following elements:

<TAXONOMY> defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.

<BIBL> contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

<CATEGORY> contains an individual descriptive category, possibly nested within a superordinate category, within a user-defined taxonomy.

<CATDESC> describes some category within a taxonomy or text typology, either in the form of a brief prose description or in terms of the situational parameters used by the TEI formal <TEXTDESC>.

In the simplest case, the taxonomy may be defined by a bibliographic reference, as in the following example:

```
<classDecl>
     <taxonomy id='LCSH'>
          <bibl>Library of Congress Subject Headings
          </bibl>
     </taxonomy>
</classDecl>
```

Alternatively, or in addition, the encoder may define a special purpose classification scheme, as in the following example:

```
<taxonomy id=B>
   <bibl>Brown Corpus</bibl>
   <category id=B.A><catDesc>Press Reportage
      <category id=B.A1><catDesc>Daily</category>
      <category id=B.A2><catDesc>Sunday</category>
      <category id=B.A3><catDesc>National</category>
      <category id=B.A4><catDesc>Provincial</category>
      <category id=B.A5><catDesc>Political</category>
      <category id=B.A6><catDesc>Sports</category>
     ...
   </category>
   <category id=B.D><catDesc>Religion
      <category id=B.D1><catDesc>Books</category>
      <category id=B.D2><catDesc>Periodicals and tracts</category>
   </category>
  ...
</taxonomy>
```

Linkage between a particular text and a category within such a taxonomy is made by means of the <CATREF> element within the <TEXTCLASS> element, as further described below.

14.3 The Profile Description

The <PROFILEDESC> element enables information characterizing various descriptive aspects of a text to be recorded within a single framework. It has three optional components:

<CREATION> contains information about the creation of a text.

<LANGUSAGE> describes the languages, sublanguages, registers, dialects etc. represented within a text.

<TEXTCLASS> groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.

Examples:

```
<creation>
     <date value='1992-08'>August 1992</date>
     <name type=place>Taos, New Mexico</name>
</creation>
```

The <TEXTCLASS> element classifies a text by reference to the system or systems defined by the <CLASSDECL> element, and contains one or or more of the following elements:

<KEYWORDS> contains a list of keywords or phrases identifying the topic or nature of a text. Attributes include:

SCHEME identifies the controlled vocabulary within which the set of keywords concerned is defined.

<CLASSCODE> contains the classification code used for this text in some standard classification system. Attributes include:

SCHEME identifies the classification system or taxonomy in use.

<CATREF> specifies one or more defined categories within some taxonomy or text typology. Attributes include:

TARGET identifies the categories concerned

The element <KEYWORDS> contains a list of keywords or phrases identifying the topic or nature of a text. The attribute SCHEME links these to the classification system defined in <TAXONOMY>.

```
<textClass>
     <keywords scheme=LCSH>
          <list>
          <item>English literature - History and criticism -
               Data processing.</item>
          <item>English literature - History and criticism -
               Theory etc.</item>
          <item>English language - Style - Data
               processing.</item>
          </list>
     </keywords>
</textClass>
```

14.4 The Revision Description

The <REVISIONDESC> element provides a change log in which each change made to a text may be recorded. The log may be recorded as a sequence of <CHANGE> elements each of which contains

<DATE> contains a date in any format.

<RESPSTMT> supplies a statement of responsibility for someone responsible for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.

<ITEM> contains one component of a list.

Example:

```
<revisionDesc>
     <change><date>6/3/91:</date>
          <respStmt><name>EMB</name><resp>ed.</resp></respStmt>
          <item>File format updated</item>
     <change><date>5/25/90:</date>
          <respSmt><name>EMB</name><resp>ed.</resp>
          <item>Stuart's corrections entered</item>
</revisionDesc>
```