TEI Tite — A recommendation for off-site text encoding

Perry Trolard, for the TEI Consortium

December 16, 2011

Version 1.1 — September 2011

1 Introduction

TEI Tite is a constrained customization of TEI designed for use when outsourcing production of TEI documents to vendors, who use some combination of OCR and keyboarding to produce encoded text. While the canonical version of Tite is maintained by the TEI Council, a derived version is used in the AccessTEI program.

TEI Tite is meant to express a transitional format for documents, not an archival one. A project outsourcing encoding of documents using Tite should convert Tite documents created by vendors into a more suitable format for long-term preservation, such as one of the encoding levels of *Best Practices for TEI in Libraries* or a project-specific TEI customization.

While Tite includes only a limited set of all of the elements in TEI, it should not be confused with *TEI Lite*, which also contains a subset of elements. What distinguishes Tite from other TEI customizations is that Tite is meant to prescribe *exactly one* way of encoding a particular feature of a document in as many cases as possible, ensuring that any two encoders would produce the same XML document for a source document.

This document specifies how a source document should be encoded using TEI Tite. Its organizing model is roughly the structure of a TEI document itself, and it proceeds from high-level features to low, starting with general requirements, text structure, directions on when to group texts, considerations about type of text (genre and format), continuing down to instructions on marking phrase-level features, reference systems, and so forth. In its original ODD (one document does-it-all) format, this document can generate everything necessary for working in TEI Tite: both documentation (this Tite-specific prose as well as the full technical documentation for each of its elements) and schemas in either W3C Schema, RELAX NG, or XML DTD. Software utilities, including the Roma web tool, can generate these.

Tite uses a subset of the TEI's elements, except for a few shortcut elements for the convenience of use by vendors (<b>, <i>, <ul>, <sup>, <sub>, <smcap>, <cols> and <ornament>) which can be transformed to normal TEI elements. Tite is also not a TEI-conformant customization since it breaks the TEI Abstract Model by omitting <teiHeader> for encoder convenience. That is, Tite was created primarily by *removing* elements and attributes from the TEI, and not from extensive *modification*. As a TEI customization, Tite inherits TEI semantics, and ambiguity in this specification should be resolved with reference to the TEI Guidelines. What makes Tite distinct is that where the TEI in general is famously tolerant of multiple methods of encoding a given feature, Tite seeks uniformity of encoding through constraint, via its stripped-down tag set and via this specification.

Tite can be used to encode printed prose, poetry, drama, newspapers, and anything else which can be described with the basic TEI building-blocks of divisions, paragraphs, line groups, and speeches.

In this documentation, *document* refers generally to the item (book, pamphlet, newspaper, etc.) to be encoded and *text* to either linguistic (as opposed to graphic) material or a logically distinct literary unit.

2 General Requirements

2.1 What to Capture

All printed material should be captured: all text (that is, printed characters) should be transcribed and the presence of graphical items or other non-transcribable elements should be indicated with markup.

2.2 End-of-line Hyphens

A distinction should be maintained in the electronic transcription between end-of-line or "soft" hyphens (an artifact of page layout) and "hard" hyphens (a linguistic feature). The former should be transcribed as the SOFT HYPHEN (U+00AD) character; the latter, as the HYPHEN-

MINUS (U+002D) character generally available on Western keyboards. In the rare case of coincidence of the two types — where a word that is normally hyphenated is split across a line break at its hyphen — the hyphen should be considered hard, and transcribed as the HYPHEN-MINUS.

2.3   Character Encoding

Characters should be encoded in UTF-8. For characters not easily input from the keyboard, use hexadecimal numeric entities (e.g. é, the small latin e with acute accent, is represented as é).

2.4   Accuracy and Verification

The standard for accuracy of transcription should be at least 99.99% (1 error in 10,000 characters). The sample size for verification will be 5% of the total text.

2.5   Documenting the Encoding Process

Almost surely, difficult encoding situations will arise whose resolution may not be covered by this documentation or the TEI Guidelines. In such cases, it is important to document the markup choices that are made. To this end each encoded file should be accompanied by a document with such notes. These notes should reference features of a document that seem remarkable to encoders and how these were handled by encoders.

3   Global Text Structure

3.1   TEI Tite text structure

In TEI Tite, <text> is the root element, containing front matter, the body of the text, and back matter.

```
<text xml:id="unique-identifier">
 <front/>
 <body/>
 <back/>
</text>
```

The <text>'s xml:id attribute should contain a unique identifier for the document being encoded.

Tite omits the <teiHeader> element as a convenience to transcribers. This departs from normal TEI practice, which requires <TEI> as the root element, containing <teiHeader> and <text> elements. In order to bring a document encoded in TEI Tite into adherence with the TEI Abstract Model, projects should add a teiHeader before engaging in post-transcription processing.

3.2   Groups of Texts

A document should be encoded as a group of texts only when each member of the group contains its own front or back matter (most often, a separate title page). In this case the <group> element should be a child of the <text> element, and should contain child <text> elements each containing a <front>, <body>, and <back> (each <text> need not have both front and back matter, but should have at least one). Note that this group of texts will still have its own front and back matter. When dealing with a group of texts, the basic TEI text structure is modified to look like:

```
<text>
 <front/>
 <group>
  <text>
   <front/>
   <body/>
   <back/>
  </text>
  <text>
   <front/>
   <body/>
   <back/>
  </text>
 </group>
 <back/>
</text>
```

In cases where a document appears to contain a group of texts but the above condition is not met, encode each unit as a (numbered) <div> with an appropriate type attribute.

3.3 Structural Divisions

Tite uses numbered divisions: <div1> through <div7>, which stand for levels of nesting within a text. <div1>s nest inside or are contained by the <front>, <body>, and <back> elements, <div2>s nest inside or are contained by <div1>s, etc. The document's table of contents is often a good place to find cues about where structural divisions start and end; other cues can be blank pages, recurring typographical or ornamental features, or a numbering system ("Chapter 5" etc.). Also, the presence of a heading will often indicate the beginning of a division.

The type attribute should be used to express the type of division being marked. Where present, use a name for division type given in the document itself. Though any constrained enumerated list of type values will have to be determined on a job-by-job basis, some examples of appropriate division types are:

- act

- article

- book

- chapter

- essay

- letter

- part

- scene

- section

- subsection

When a heading is present, encode it with the <head> element. If there is more than one heading at the beginning of a given division, encode each heading with its own <head> element, using the type attribute to distinguish them. Appropriate values are:

- main

- sub (subtitle)

- alt (alternate)

- desc (descriptive)

The n attribute should be used to record sequential labels associated with a structural division (numbers, numerals, letters). When present, these labels should also be transcribed within the content of <head> element. For instance:

```
<div1 n="III" type="part">
 <head>III: It Awakes</head>
</div1>
```

3.3.1 False Indicators

A *divisional title* is a page that resembles a half-title page: it displays the title or heading of a major structural unit on an otherwise blank page. *Divisional titles* should be encoded not with a separate <div> element, but as a <head> within the appropriate <div>. For *half-title pages* and similar *fly-title pages* see the section on Front Matter.

Another potential false indication of a new structural division is an *ornament* used as an informal division: a printer's ornament of some sort, a string of asterisks or periods, or a horizontal line. Mark these with the special <ornament> element. If the ornament is a horizontal line or printer's device or otherwise not transcribable, make the element empty and include an appropriate type attribute (line or ornament); if the ornament is made up of characters, transcribe the characters into the <ornament>'s content.

3.4 Front and Back Matter

Front and back matter should be encoded with the <front> and <back> elements, respectively. <div1> elements should contain the major sections and should be characterized by type attribute values. The exception, however, is the title page, which should be encoded with the <titlePage> element and its children. The <titlePart> element should have a type attribute with one of the following values:

- main

- sub (subtitle)

- desc (descriptive title)

- alt (alternate title)

- volume (volume information)

<titlePart type="volume"> should be used to encode volume information wherever it is found on the title page, even if it is separated from the other title information. The elements that make up the <titlePage> content model are: <graphic>, <byline>, <epigraph>, <docTitle>, <titlePart>, <docAuthor>, <docEdition>, <docImprint>, <docDate>, <figure>, <ornament>.

Information on the verso of the title page should be included as well (after a <pb>).

Common items to encode in front and back matter – and therefore common type attribute values for <front> and <back> divisions are: **front**

- acknowledgements

- advertisement

- castlist

- contents

- dedication

- fly-title

- foreword

- introduction

- preface

back

- appendix

- bibliography

- colophon

- glossary

- index

Half-title and *fly-title* pages may be encountered in the front matter. A *half-title* page precedes the title page proper and sometimes includes volume or series information; a *fly-title* page comes at the very end of the front matter, just before the body. In the case of half-titles, encode these as <div1 type="half-title"> (with <titlePart> elements as appropriate); in the case of fly-titles, encode them likewise with <div1 type="fly-title">, making sure to make the fly-title division the last part of the front matter (and not the first part of the body, as may seem reasonable as well).

4 Types of Text

Tite is equipped to support basic encoding of several types of text: in terms of genre, it supports prose, verse, and drama, and in terms of format, it supports books, newspapers, pamphlets, and other similar printed material. Tite has special elements for letters, verse, drama, and newspapers.

4.1 Letters

<opener> and <closer> are elements designed to encode the beginning and ending sections of letters, prefaces, diary entries, or other personal types of writing. Both elements contain:

- <dateline>: for recording time and place of composition; use <date> with when value (formatted yyyy-mm-dd) to record date information

- <signed>: for recording a signature

- <salute>: for recording salutation at the beginning ("Dear Roger,") or end ("Yours truly,")

<opener> contains the additional elements <epigraph>, <argument>, and <byline>. <epigraph> will often be useful in the context of a letter. When encoding an epigraph, make sure to encode the content as you would any other feature, marking line groups, bibliographical elements, etc.

<argument> and <byline>, however, are not intended specifically for use with letters:

- <argument>: for a summary that precedes a division

- <byline>: for a statement of responsibility for the document

4.2   Verse

All verse should be encoded within at least one <lg> element, even when there are no distinct stanzas or when the verse is interspersed with prose. If it is known, use the type attribute to express the type of line group. Sometimes within a poem there is a question about what should be tagged as a <lg> or as a separate <div>. As a rough rule of thumb, if there is a title accompanying the division, use the <div> element; otherwise, use <lg>.

Each line of verse should be encoded with the <l> element, and care should be taken to distinguish these logical lines of verse from lines motivated by page layout. The latter should be encoded as <lb>s. Thus

```
AS virtuous men pass mildly away,
    And whisper to their souls to
go,
Whilst some of their sad friends
do say,
    "Now his breath goes," and
some say, "No."
```

should be encoded as

```
<lg type="stanza">
 <l>AS virtuous men pass mildly away,</l>
 <l rend="indent(1)">And whisper to their souls to<lb/> go,</l>
 <l>Whilst some of their sad friends<lb/> do say,</l>
 <l rend="indent(1)">"Now his breath goes," and<lb/> some say, "No."</l>
</lg>
```

Also, as in the example above, use the rend attribute to mark when a line is indented more than its siblings. Use "numbered" indent values (e.g. "indent(1)", "indent(2)", etc.) to make clear levels of indentation.

4.3   Drama

The standard TEI elements for drama should be used: <sp>, <stage>, <speaker>. If the who attribute is used on <sp>, also transcribe who is given as the speaker, in whatever form it is written, in the <speaker> element. Short pieces of stage direction that accompany the speaker designation may be included in the <speaker> element.

Scenes and acts should be encoded as appropriately nested <div> elements with type attributes of scene or act, respectively. Cast lists can likewise be encoded using <div> and `type="castlist"`.

Prologues and epilogues can be treated as <sp>s of their own, unless their structure would be better represented by nested <div> elements.

4.4   Newspapers

Tite includes the elements <cols> and <cb> which are well suited for the multi-column layout of newspapers. Additional relevant elements are: <ref>, to encode a pointer to the continuation of a story in a different column or on a different page; and <figure>, to describe illustrations, advertisements, and cartoons.

5 Block-level Features

5.1 Block Quotations

Use the <q> element to encode block quotations. A block quotation is indicated by its being set off from surrounding text either with extra line-spacing or margins or with a different typeface. If the quotation is of an entire text, use the <floatingText> element and its children inside the <q> element:

```
<div1 type="intro">
 <p/>
 <q>
  <floatingText>
   <body>
    <lg type="poem"/>
   </body>
  </floatingText>
 </q>
 <p/>
</div1>
```

If present, transcribe all quotation marks or other delimiters inside the <q> element.

5.2 Figures

Use the <figure> element to encode figures. If a figure has a heading or caption, encode it with the <head> element. If there is associated text, simply use a <p> to encode it.

5.3 Tables and Lists

Tables and lists are encoded as in the TEI Guidelines, but note the following.

If a cell in a table is a heading or a label, set the role attribute to label; if the cell contains data, there is no need to use role: data is the default. If a cell or row spans more than one column or row, use the rows or cols attributes set to the number of columns or rows that it spans.

If unsure about whether a structure is best encoded as a list or table, record it as a table only if it would not be properly understood without tabular layout.

Lists should be encoded as either sequences of <items> or <label>-<item> pairs. When items in the list contain a label, as in a gloss list, be sure to use the latter form.

5.4 Notes

Both the reference to the note in the running text and the note itself must be encoded. Use <ptr> or <ref> to encode the reference. If there is no reference in the text (often the case for marginal notes), supply a <ptr> element in a reasonable place in the text running beside the note. If there is a reference (number, symbol, etc.), use the <ref> element and include the reference text as the content. In both cases, a target attribute must be supplied which contains the xml:id value of the associated <note>.

When encoding the note itself with the <note> element, the xml:id and place attributes must be supplied. See the TEI documentation for acceptable values for place; the most common will be foot, end, margin-left (-right, -top, -bot).

Transcribe the note directly after it is referenced in the document. In the case of notes without explicit reference (pointed to with <ptr>), set the anchored attribute to false.

5.5 "divWrapper" Elements

Elements that can appear at the beginning and end of structural divisions, such as <argument>, <epigraph>, and <opener>, are called "divWrapper" elements in the TEI class system. An

argument is a summary of what is to come; be sure to distinguish this from a *heading*, which is a *title* for the division. If an epigraph comes with bibliographic or simple citation material, encode this as well. For example:

```
<epigraph>
 <cit>
  <q>"I have sworn upon the altar of God eternal hostility against every form of
tyranny over
     the mind of man."</q>
  <bibl>
   <author>Thomas Jefferson.</author>
  </bibl>
 </cit>
</epigraph>
```

5.6   Uncertain Blocks

In rare cases where the logical identity of a block-level element is hard to discern, use the TEI element <ab> (anonymous block) instead of applying a <p> or <div> element. In these cases, be sure to document this decision in accompanying notes. *Applying this element should be viewed as a last resort.*

The <gap> element should be used when for some reason the document being transcribed contains illegible text (smudged, torn, missing, etc.) or something outside the scope of transcription for a given project: characters in an unsupported character set, for instance. <gap> indicates that something is omitted. When using <gap>, set the reason attribute to an appropriate value. (See <unclear> below.)

6   Phrase-level Features

6.1   Typographical Changes

There are six elements in Tite that capture specific typographical features:

<b>   for bold-face glyphs

<i>   for italicized glyphs

<ul>   for underlined glyphs

<smcap>   for glyphs in small-caps

<sub>   for glyphs in subscript

<sup>   for glyphs in superscript

These mark the physical change, and are agnostic about a logical motivation for it. There are two exceptions to this approach, however: marking foreign words and titles. In the case of foreign words, use the <foreign> element; in the case of titles, use the <title> element only if certain that the word or phrase in question is a title. If a phrase is, say, italicized, but you are uncertain about its being a title, use the <i> element instead. Foreign words should be marked only if they are typographically distinguished from surrounding text.

If there is a typographical feature not covered by the above elements, the TEI <hi> element is still available in Tite. Use it without a rend attribute.

6.2   Phrase-level Quotation

For passages set off by quotation marks or another delimeter, use the <q> element, including the delimeter inside the tag.

6.3 Alignment and Indentation

If the alignment of an element seems remarkable, set the element's rend attribute to an appropriate value (normally center, right, left, etc.). However, when semantic already accounts for its cause, description of alignment is not necessary. Headings, for instance, do not need to be marked as being centered.

To indicate level of indentation (often in verse), use numerical "arguments" to "indent", as in indent(1), indent(-1), and so on.

6.4 Uncertain Segments

The <seg> element is the phrase-level analogue to the <ab> element. If a phrase-level feature seems to be present but its identity is hard to fathom, use this element. This, again, is a last resort.

Alternately, when a passage of text is for some reason too hard to read, use the <unclear> element, setting the reason attribute to an appropriate value. When using <unclear>, surround the entire word with the tag if any part of it is unclear (not just the illegible letter, say).

6.5 Unknown Glyphs

For cases in which it is unknown which character a given glyph corresponds to, mark the glyph with the <g> element to indicate the uncertainty. By convention in Tite, <g> represents any unknown glyph; no ref attribute is necessary. Note that unknown glyphs are different from *illegible* text.

7 Reference Systems

Encode page breaks (<pb>) at the *start* of each page, and encode breaks even for blank pages. If the page is numbered, include the page number as the value of the n attribute and, again, no matter where the page number is printed on the page, place the <pb> element at the "top."

If marking column breaks, follow the same rules as for page breaks. Column breaks are imagined to appear at the *top* of the column, at the beginning of the column's text. The <cols> element exists to record a change in columnar layout. If such a change occurs, mark the beginning of the new layout with <cols> and supply the new number of columns as the value for the n attribute.

If line breaks are to be captured, use the <lb> element.

Appendices

TEI Tite and the *Best Practices for TEI in Libraries*

The *Best Practices for TEI in Libraries* ("BP") creates common definitions of "levels of encoding" based on depth of markup applied. Because the levels of encoding provide a tremendously useful common set of terms, it's helpful to situate TEI Tite according to them.

Mapped to BP levels, TEI Tite would sit between Level 3 and Level 4: it requires use of all the elements from Level 3 plus additional ones, but requires fewer elements than Level 4. Relative to Level 3, "Simple Analysis," Tite

- encourages the use of the rend attribute on typographically distinct text (marked with <hi>), implicitly, through the provision of convenience elements (<i>, <b>, etc.), and it provides the <title> and <foreign> elements for semantic markup of typographically distinct phrases; in level 3, the rend attribute is optional, and <title> and <foreign> are not provided

- provides some genre-specific elements in addition to those for verse that level three also provides (<lg>, <l>): <sp>, <speaker>, and <stage> for drama, the <cols> element especially for newspapers.

The most useful comparison for Tite is to Level 4 ("Basic Content Analysis"), provides the most useful comparison. The folowing items represent instances where Tite is *less* ambitious than Level 4:

- except in the case of the <foreign> and <title> elements, it is preferred in Tite to describe typographical changes physically, rather than semantically; Tite uses <i>, <b>, etc. where level four uses <emph>, <gloss>, <term>

- Tite provides only <q> for quoted material, where level four is more discriminating, using <quote>, <said>, <mentioned>, <soCalled>

- Tite doesn't provide elements for editorial intervention, as level four does: <choice>, <sic>, <corr>

- Tite doesn't provide entity-specific naming elements, like <persName>, <placeName>, <orgName> and their list- (<listPerson>, etc.) forms

Bringing Tite-encoded documents up to BP Level 4 would simply require application of additional markup, not significant reworking of markup, and in that way Tite is compatible with the BP.

Do also keep in mind that Tite lacks both the <teiHeader> and root <TEI> element used in TEI-conformant documents.

Formal specification

Schema tei_tite: changed components

<ab> (anonymous block) contains any arbitrary component-level unit of text, acting as an anonymous container for phrase or inter level elements analogous to, but without the semantic baggage of, a paragraph.16.3.

Module linking

Used by model.pLike

Contained by

core: item note q sp stage

figures: figure

textstructure: argument body div1 div2 div3 div4 div5 div6 div7 epigraph postscript

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list
        listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration   `element ab { macro.paraContent }`

Example

```
    <div type="book" n="Genesis">
     <div type="chapter" n="1">
      <ab>In the beginning God created the heaven and the earth.</ab>
      <ab>And the earth was without form, and void; and
          darkness was upon the face of the deep. And the
          spirit of God moved upon the face of the waters.</ab>
      <ab>And God said, Let there be light: and there was light.</ab>
     </div>
    </div>
```

Note   The <ab> element may be used at the encoder's discretion to mark any
component-level elements in a text for which no other more specific appropriate
markup is defined.

<abbr> (abbreviation) contains an abbreviation of any sort.3.5.5.

Module   core

Attributes   Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by   model.choicePart model.pPart.editorial

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l
        label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
        salute signed titlePart

May contain

core: abbr add address cb date del email foreign gap graphic hi lb milestone name note
        num pb ptr ref time title unclear

derived-module-tei_tite: b cols i smcap sub sup ul

figures: figure formula

gaiji: g

linking: seg

transcr: handShift

Declaration

```
element abbr { att.global.attributes, macro.phraseSeq }
```

Example

```
<choice>
 <expan>North Atlantic Treaty Organization</expan>
 <abbr cert="low">NorATO</abbr>
 <abbr cert="high">NATO</abbr>
 <abbr cert="high" xml:lang="fr">OTAN</abbr>
</choice>
```

Example

```
<choice>
 <abbr>SPQR</abbr>
 <expan>senatus populusque romanorum</expan>
</choice>
```

Note The <abbr> tag is not required; if appropriate, the encoder may transcribe abbreviations in the source text silently, without tagging them. If abbreviations are not transcribed directly but *expanded* silently, then the TEI header should so indicate.

---

att.datable.w3c provides attributes for normalization of elements that contain datable events using the W3C datatypes.

Module tei

Members att.datable [date time]

Attributes Attributes

@when supplies the value of the date or time in a standard form, e.g. yyyy-mm-dd.

Status Optional

Datatype `data.temporal.w3c`

Values A normalized form of temporal expression conforming to the W3C *XML Schema Part 2: Datatypes Second Edition.*
Examples of W3C date, time, and date & time formats.

```
<p>
 <date
   when="1945-10-24">24 Oct 45</date>
 <date
   when="1996-09-24T07:25:00Z">September 24th, 1996 at 3:25
in the morning</date>
 <time
   when="1999-01-04T20:42:00-05:00">Jan 4 1999 at 8
pm</time>
 <time
   when="14:12:38">fourteen twelve and 38 seconds</time>
 <date
   when="1962-10">October of 1962</date>
 <date
   when="--06-12">June 12th</date>
 <date
   when="---01">the first of the month</date>
```

```
    <date
      when="--08">August</date>
    <date
      when="2006">MMVI</date>
    <date
      when="0056">AD 56</date>
    <date
      when="-0056">56 BC</date>
</p>
```

```
This list begins in
the year 1632, more precisely on Trinity Sunday, i.e. the
Sunday after
Pentecost, in that year the <date
  calendar="#Julian"
  when="1632-06-06">27th of May (old style)</date>.
```

```
<opener>
 <dateline>
  <placeName>Dorchester, Village,</placeName>
  <date
    when="1828-03-02">March 2d. 1828.</date>
 </dateline>
 <salute>To
   Mrs. Cornell,</salute> Sunday <time
   when="12:00:00">noon.</time>
</opener>
```

Note The value of the when attribute should be the normalized representation of the date, time, or combined date & time intended, in any of the standard formats specified by *XML Schema Part 2: Datatypes Second Edition*, using the Gregorian calendar. The most commonly-encountered format for the date part of the when attribute is `yyyy-mm-dd`, but `yyyy`, `--mm`, `---dd`, `yyyy-mm`, or `--mm-dd` may also be used. For the time part, the form `hh:mm:ss` is used. Note that this format does not currently permit use of the value 0000 to represent the year 1 BCE; instead the value -0001 should be used.

@from indicates the starting point of the period in standard form, e.g. yyyy-mm-dd.
  Status Optional

  Datatype | data.temporal.w3c |

  Values A normalized form of temporal expression conforming to the W3C *XML Schema Part 2: Datatypes Second Edition.*

@to indicates the ending point of the period in standard form, e.g. yyyy-mm-dd.
  Status Optional

  Datatype | data.temporal.w3c |

  Values A normalized form of temporal expression conforming to the W3C *XML Schema Part 2: Datatypes Second Edition.*

att.global provides attributes common to all elements in the TEI encoding scheme.

Module tei

Members p foreign hi q desc gap add del unclear name email address addrLine num date time abbr ptr ref list item label head note graphic milestone pb lb cb author editor respStmt resp title publisher pubPlace bibl listBibl l lg sp speaker stage text body group floatingText div1 div2 div3 div4 div5 div6 div7 byline dateline argument epigraph opener closer salute signed postscript titlePage docTitle titlePart docAuthor docEdition docImprint docDate front back table row cell formula figure seg g handShift b i ul sub sup smcap cols ornament

Attributes Attributes

@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

Status Optional

Datatype `xsd:ID`

Values any valid XML identifier.

Note The xml:id attribute may be used to specify a canonical reference for an element; see section 3.10. .

@n (number) gives a number (or other label) for an element, which is not necessarily unique within the document.

Status Optional

Datatype 1–∞ occurrences of `data.word` separated by whitespace

Values the value consists of one or more instances of data.word. Each instance may contain only letters, digits, punctuation characters, or symbols: it may not contain whitespace or word separating characters. It need not be restricted to numbers.

Note The n attribute may be used to specify the numbering of chapters, sections, list items, etc.; it may also be used in the specification of a standard reference system for the text.

@xml:lang (language) indicates the language of the element content using a "tag" generated according to BCP 47

Status Optional

Datatype `data.language`

Values The value must conform to BCP 47. If the value is a private use code (i.e., starts with `x-` or contains `-x-`) it should, and if not it may, match the value of an ident attribute of a <language> element supplied in the TEI Header of the current document.

Note the xml:lang value will be inherited from the immediately enclosing element, or from its parent, and so on up the document hierarchy. It is generally good practice to specify xml:lang at the highest appropriate level, noticing that a different default may be needed for the teiHeader from that needed for the associated resource element or elements, and that a single TEI document may contain texts in many languages.

@rend (rendition) indicates how the element in question was rendered or presented in the source text.

Status Optional

Datatype 1–∞ occurrences of `data.word` separated by whitespace

Values may contain any number of tokens, each of which may contain

letters, punctuation marks, or symbols, but not word-separating characters.

```
<head
  rend="align(center) case(allcaps)">
 <lb/>To The <lb/>Duchesse <lb/>of <lb/>Newcastle,
<lb/>On Her <lb/>
 <hi
    rend="case(mixed)">New Blazing-World</hi>.
</head>
```

Note These Guidelines make no binding recommendations for the values of the rend attribute; the characteristics of visual presentation vary too much from text to text and the decision to record or ignore individual characteristics varies too much from project to project. Some potentially useful conventions are noted from time to time at appropriate points in the Guidelines.

---

att.pointing defines a set of attributes used by all elements which point to other elements by means of one or more URI references.

Module tei

Members ptr ref note

Attributes Attributes

@target specifies the destination of the reference by supplying one or more URI References

Status Optional

Datatype 1–∞ occurrences of $\boxed{\texttt{data.pointer}}$ separated by whitespace

Values One or more syntactically valid URI references, separated by whitespace. Because whitespace is used to separate URIs, no whitespace is permitted inside a single URI. If a whitespace character is required in a URI, it should be escaped with the normal mechanism, e.g. `TEI%20Consortium`.

---

att.typed provides attributes which can be used to classify or subclassify elements in any way.

Module tei

Members add del name date time ptr ref head note milestone pb lb cb bibl listBibl lg text floatingText div1 div2 div3 div4 div5 div6 div7 figure seg g

Attributes Attributes

@type characterizes the element in some sense, using any convenient classification scheme or typology.

Status Optional

Datatype $\boxed{\texttt{data.enumerated}}$

Note The type attribute is present on a number of elements, not all of which are members of `att.typed`.

<b> (bold) for capturing typographical feature: bold glyphs.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.hiLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element b { att.global.attributes, macro.paraContent }
```

<cit> (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example. 3.3.3. 4.3.1. 9.3.5.1.

Module core

Used by model.quoteLike

Contained by

core: add cit del desc head hi item l note p q ref sp stage title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure

linking: ab seg

textstructure: argument body div1 div2 div3 div4 div5 div6 div7 docEdition epigraph postscript titlePart

May contain

core: bibl cb cit gap lb milestone note pb ptr q ref

derived-module-tei_tite: cols

figures: figure

textstructure: floatingText

Declaration

```
element cit
{
   (
      model.qLike    | model.egLike    | model.biblLike
      | model.ptrLike    | model.global    | model.entryPart   )+
}
```

Example

```
<cit>
 <quote>and the breath of the whale is frequently attended with such an
insupportable smell,
   as to bring on disorder of the brain.</quote>
 <bibl>Ulloa's South America</bibl>
</cit>
```

Example

```
<entry>
 <form>
  <orth>horrifier</orth>
 </form>
 <cit type="translation" xml:lang="en">
  <quote>to horrify</quote>
 </cit>
 <cit type="example">
  <quote>elle était horrifiée par la dépense</quote>
  <cit type="translation" xml:lang="en">
   <quote>she was horrified at the expense.</quote>
  </cit>
 </cit>
</entry>
```

<cols/> (columns) with the "n" attribute (denoting new number of columns) is used to mark where a document changes columnar layout.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

> @ed indicates the edition or version in which the change in columnar layout is located at this point
>
> > *Status* Optional
> >
> > *Datatype* `data.code`

Used by model.milestoneLike

Contained by

core: abbr add addrLine address author bibl cit date del editor email foreign head hi item l label lg list listBibl name note num p pubPlace publisher q ref resp sp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure table

linking: ab seg

textstructure: argument back body byline closer dateline div1 div2 div3 div4 div5 div6 div7 docAuthor docDate docEdition docImprint docTitle epigraph floatingText front group opener postscript salute signed text titlePage titlePart

May contain Empty element

Declaration

```
element cols
{
   att.global.attributes,
   attribute [http://www.tei-c.org/ns/tite/1.0]ed { data.code }?,
   empty
}
```

<date> contains a date in any format.3.5.4. 2.2.4. 2.5. 3.11.2.3. 15.2.3. 13.3.6.

Module core

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*) att.datable (*@calendar, @period*) (att.datable.w3c (*@when, @from, @to*)) att.typed (*@type*)

Used by model.dateLike model.publicationStmtPart

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address cb date del email foreign gap graphic hi lb milestone name note num pb ptr ref time title unclear

derived-module-tei_tite: b cols i smcap sub sup ul

figures: figure formula

gaiji: g

linking: seg

transcr: handShift

Declaration

```
element date
{
   att.global.attributes,
   att.datable.attribute.period,
   att.datable.w3c.attributes,
   att.typed.attributes,
   ( text | model.gLike | model.phrase | model.global )*
}
```

Example

```
<date when="1980-02">early February 1980</date>
```

Example

```
Given on the <date when="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>
```

Example

```
<date when="1990-09">September 1990</date>
```

\<formula\> contains a mathematical or other formula.14.2.

Module figures

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.graphicLike

Contained by

core: abbr add addrLine author date del editor email foreign head hi item l label name
note num p pubPlace publisher q ref speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure formula table

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
salute signed titlePart

May contain

core: graphic

figures: formula

Declaration

```
element formula { att.global.attributes, ( text | model.graphicLike )* }
```

Example

```
<formula notation="tex">$e=mc^2$</formula>
```

Note

\<g\> (character or glyph) represents a non-standard character or glyph.

Module gaiji

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*) att.typed (*@type*)

Used by model.gLike

Contained by

core: abbr add addrLine author bibl date del editor email foreign head hi item l label
name note num p pubPlace publisher q ref speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
salute signed titlePart

May contain Character data only

Declaration

```
element g { att.global.attributes, att.typed.attributes, text }
```

Example

```
<g ref="#ctlig">ct</g>
```

This example points to a <glyph> element with the identifier `ctlig` like the following:

```
<glyph xml:id="ctlig"/>
```

Example

```
<g ref="#per-glyph">per</g>
```

The medieval brevigraph per could similarly be considered as an individual glyph, defined in a <glyph> element with the identifier `per` like the following:

```
<glyph xml:id="per-glyph"/>
```

Note The name `g` is short for *gaiji*, which is the Japanese term for a non-standardized character or glyph.

<gap>

<**gap**> (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, invisible, or inaudible.3.4.3.

Module core

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*)

    @reason gives the reason for omission. Sample values include sampling, inaudible, irrelevant, cancelled.

        Status Optional

        Datatype 1–∞ occurrences of `data.word` separated by whitespace

        Values any short indication of the reason for the omission.

Used by model.global.edit

Contained by

core: abbr add addrLine address author bibl cit date del editor email foreign head hi item l label lg list name note num p pubPlace publisher q ref resp sp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure table

linking: ab seg

textstructure: argument back body byline closer dateline div1 div2 div3 div4 div5 div6 div7 docAuthor docDate docEdition docImprint docTitle epigraph floatingText front group opener postscript salute signed text titlePage titlePart

May contain

core: desc

Declaration

```
element gap
{
   attribute reason { list { data.word, data.word* } }?,
```

```
        att.global.attributes,
        model.glossLike*
    }
```

Example

```
    <gap extent="4" unit="chars" reason="illegible"/>
```

Example

```
    <gap extent="1" unit="essay" reason="sampling"/>
```

Example

```
    <del>
     <gap
        atLeast="4"
        atMost="8"
        unit="chars"
        reason="illegible"/>
    </del>
```

Note The <gap>, <unclear>, and <del> core tag elements may be closely allied in use with the <damage> and <supplied> elements, available when using the additional tagset for transcription of primary sources. See section 11.3.3.2. for discussion of which element is appropriate for which circumstance. The <gap> tag simply signals the editors decision to omit or inability to transcribe a span of text. Other information, such as the interpretation that text was deliberately erased or covered, should be indicated using the relevant tags, such as <del> in the case of deliberate deletion.

<graphic> indicates the location of an inline graphic, illustration, or figure.3.9.

Module core

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

    @url (uniform resource locator) A URL which refers to the image itself.

        Status Required

        Datatype `data.pointer`

Used by model.graphicLike model.titlepagePart

Contained by

core: abbr add addrLine author date del editor email foreign head hi item l label name note num p pubPlace publisher q ref speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure formula table

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePage titlePart

May contain

core: desc

Declaration

```
element graphic
{
    attribute url { data.pointer },
```

```
       att.global.attributes,
       model.glossLike*
}
```

Example

```
  <figure>
   <graphic url="fig1.png"/>
   <head>Figure One: The View from the Bridge</head>
   <figDesc>A Whistleresque view showing four or five sailing boats in the
   foreground, and a
      series of buoys strung out between them.</figDesc>
  </figure>
```

Note The mimeType attribute should be used to supply the MIME media type of the
image specified by the url attribute.

\<i\> (italics) for capturing typographical feature: italicized glyphs.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.hiLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l
label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list
listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element i { att.global.attributes, macro.paraContent }
```

\<l\> (verse line) contains a single, possibly incomplete, line of verse.3.12.1. 3.12. 7.2.5.

Module core

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.lLike

Contained by

core: head item lg note q sp stage

figures: figure

textstructure: argument body div1 div2 div3 div4 div5 div6 div7 epigraph postscript

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list
       listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element l { att.global.attributes, macro.paraContent }
```

Example

```
<l met="-/-/-/-/-/" part="Y"/>
```

\<note\>   contains a note or annotation.3.8.1. 2.2.6. 3.11.2.6. 9.3.5.4.

Module   core

Attributes   Attributes att.global (*@xml:id, @n, @xml:lang, @rend*) att.pointing (*@target*)
       att.typed (*@type*)

Used by   model.noteLike

Contained by

core: abbr add addrLine address author bibl cit date del editor email foreign head hi
       item l label lg list name note num p pubPlace publisher q ref resp sp speaker stage
       time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure table

linking: ab seg

textstructure: argument back body byline closer dateline div1 div2 div3 div4 div5 div6
       div7 docAuthor docDate docEdition docImprint docTitle epigraph floatingText front
       group opener postscript salute signed text titlePage titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi l label lb lg
       list listBibl milestone name note num p pb ptr q ref sp stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: ab seg

textstructure: floatingText

transcr: handShift

Declaration

```
element note
{
   att.global.attributes,
   att.pointing.attributes,
   att.typed.attributes,
   macro.specialPara}
```

Example In the following example, the translator has supplied a footnote containing an explanation of the term translated as "painterly":

```
And yet it is not only
in the great line of Italian renaissance art, but even in the
painterly <note place="bottom" type="gloss" resp="#MDMH">
 <term xml:lang="de">Malerisch</term>. This word has, in the German, two
distinct meanings, one objective, a quality residing in the object,
the other subjective, a mode of apprehension and creation. To avoid
confusion, they have been distinguished in English as
<mentioned>picturesque</mentioned> and
<mentioned>painterly</mentioned> respectively.
</note> style of the
Dutch genre painters of the seventeenth century that drapery has this
psychological significance.
```

For this example to be valid, the code MDMH must be defined elsewhere, for example by means of a responsibility statement in the associated TEI Header:

```
<respStmt xml:id="MDMH">
 <resp>translation from German to English</resp>
 <name>Hottinger, Marie Donald Mackie</name>
</respStmt>
```

Example The global n attribute may be used to supply the symbol or number used to mark the note's point of attachment in the source text, as in the following example:

```
Mevorakh b. Saadya's mother, the matriarch of the
family during the second half of the eleventh century,
<note n="126" anchored="true"> The
alleged mention of Judah Nagid's mother in a letter from 1071 is, in fact,
a reference to
Judah's children; cf. above, nn. 111 and 54. </note> is well known from
Geniza documents
published by Jacob Mann.
```

However, if notes are numbered in sequence and their numbering can be reconstructed automatically by processing software, it may well be considered unnecessary to record the note numbers.

<**num**> (number) contains a number, written in any form.3.5.3.

Module core

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

@type indicates the type of numeric value.

Status Optional

Datatype `data.enumerated`

Suggested values include: **cardinal** absolute number, e.g. 21, 21.5

ordinal ordinal number, e.g. 21st

fraction fraction, e.g. one half or three-quarters

percentage a percentage

Note If a different typology is desired, other values can be used for this attribute.

Used by model.measureLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address cb date del email foreign gap graphic hi lb milestone name note num pb ptr ref time title unclear

derived-module-tei_tite: b cols i smcap sub sup ul

figures: figure formula

gaiji: g

linking: seg

transcr: handShift

Declaration

```
element num
{
   attribute type
   {
      "cardinal" | "ordinal" | "fraction" | "percentage" | xsd:Name
   }?,
   att.global.attributes,
   macro.phraseSeq}
```

Example

```
<p>I reached <num type="cardinal" value="21">twenty-one</num> on
my <num type="ordinal" value="21">twenty-first</num> birthday</p>
<p>Light travels at <num value="3E10">3×10<hi rend="sup">10</hi>
 </num> cm per second.</p>
```

Note Detailed analyses of quantities and units of measure in historical documents may also use the feature structure mechanism described in chapter 18. . The <num> element is intended for use in simple applications.

<ornament> for capturing typographical feature: printer's ornament, horizontal line, strings of asterisks or periods, etc, indicating an informal division that does not call for a new <div> element. If a horizontal rule or printer's ornament, use appropriate rend attribute and leave the element empy; if the ornament can be represented with characters, include these in the element.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*)

Used by model.inter model.titlepagePart

Contained by

core: add del desc head hi item l note p q ref stage title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure

linking: ab seg

textstructure: argument body div1 div2 div3 div4 div5 div6 div7 docEdition epigraph
     postscript titlePage titlePart

May contain Character data only

Declaration | `element ornament { att.global.attributes, text }` |

<pb/> (page break) marks the boundary between one page of a text and the next in a standard reference system.3.10.3.

Module core

Attributes Attributes att.global.facs (*@facs*) att.global (*@xml:id, @n, @xml:lang, @rend*)
     att.typed (*@type*)

Used by model.milestoneLike

Contained by

core: abbr add addrLine address author bibl cit date del editor email foreign head hi
     item l label lg list listBibl name note num p pubPlace publisher q ref resp sp speaker
     stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure table

linking: ab seg

textstructure: argument back body byline closer dateline div1 div2 div3 div4 div5 div6
     div7 docAuthor docDate docEdition docImprint docTitle epigraph floatingText front
     group opener postscript salute signed text titlePage titlePart

May contain Empty element

Declaration

```
element pb
{
   att.global.facs.attributes,
   att.global.attributes,
   att.typed.attributes,
   empty
}
```

Example Page numbers may vary in different editions of a text.

```
<p> ... <pb n="145" ed="ed2"/>
... <pb n="283" ed="ed1"/>
... </p>
```

Example A page break may be associated with a facsimile image of the page it introduces
     by means of the facs attribute

```
<body>
 <pb n="1" facs="page1.png"/>
 <p/>
 <pb n="2" facs="page2.png"/>
```

```
<p/>
</body>
```

Note By convention, <pb> elements should appear at the start of the page to which they refer. The global n attribute indicates the number or other value associated with the page which follows. This will normally be the page number or signature printed on it, since the physical sequence number is implicit in the presence of the <pb> element itself. The type attribute may be used to characterize the page break in any respect, for example as word-breaking or not.

<ptr/> (pointer) defines a pointer to another location.3.6. 16.1.

Module core

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*) att.pointing (*@target*) att.typed (*@type*)

Used by model.ptrLike

Contained by

core: abbr add addrLine author bibl cit date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain Empty element

Declaration

```
element ptr
{
   att.global.attributes,
   att.pointing.attributes,
   att.typed.attributes,
   empty
}
```

Schematron <sch:report test="@target and @cRef">Only one of the attributes 'target' and 'cRef' may be supplied.</sch:report>

Example

```
<ptr target="#p143 #p144"/>
<ptr target="http://www.tei-c.org"/>
```

<ref> (reference) defines a reference to another location, possibly modified by additional text or comment.3.6. 16.1.

Module core

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*) att.pointing (*@target*) att.typed (*@type*)

Used by model.ptrLike

Contained by

core: abbr add addrLine author bibl cit date del desc editor email foreign head hi item l
    label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei__tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
    salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list
    listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei__tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element ref
{
   att.global.attributes,
   att.pointing.attributes,
   att.typed.attributes,
   macro.paraContent}
```

Schematron <sch:report test="@target and @cRef">Only one of the attributes 'target'
    and 'cRef' may be supplied.</sch:report>

Example

```
<ref
  target="http://www.natcorp.ox.ac.uk/Texts/A02.xml#s2"> See especially the
second
sentence</ref> See also <ref>s.v. <term>locution</term>
</ref>.
```

Note The target and cRef attributes are mutually exclusive.

<smcap> (smallcaps) for capturing typographical feature: glyphs in small capitals.

Module derived-module-tei__tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.hiLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l
    label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei__tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
    salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list
      listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element smcap { att.global.attributes, macro.paraContent }
```

<stage> (stage direction) contains any kind of stage direction within a dramatic text or fragment.3.12.2. 3.12. 7.2.4.

Module   core

Attributes   Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by   model.stageLike

Contained by

core: add del desc head hi item l lg note p q ref sp stage title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell figure

linking: ab seg

textstructure: argument body div1 div2 div3 div4 div5 div6 div7 docEdition epigraph
      postscript titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi l label lb lg
      list listBibl milestone name note num p pb ptr q ref sp stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: ab seg

textstructure: floatingText

transcr: handShift

Declaration

```
element stage { att.global.attributes, macro.specialPara }
```

Example

```
<stage type="setting">A curtain being drawn.</stage>
<stage type="setting">Music</stage>
<stage type="entrance">Enter Husband as being thrown off his horse.</stage>
<stage type="exit">Exit pursued by a bear.</stage>
<stage type="business">He quickly takes the stone out.</stage>
<stage type="delivery">To Lussurioso.</stage>
<stage type="novelistic">Having had enough, and embarrassed for the
family.</stage>
<stage type="modifier">Disguised as Ansaldo.</stage>
```

```
    <stage type="location">At a window.</stage>
    <stage rend="inline" type="delivery">Aside.</stage>
```

<sub> (subscript) for capturing typographical feature: subscript glyphs.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.hiLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element sub { att.global.attributes, macro.paraContent }
```

<sup> (superscript) for capturing typographical feature: superscript glyphs.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.hiLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei__tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element sup { att.global.attributes, macro.paraContent }
```

\<time\> contains a phrase defining a time of day in any format.3.5.4.

Module core

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*) att.datable (*@calendar, @period*) (att.datable.w3c (*@when, @from, @to*)) att.typed (*@type*)

Used by model.dateLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei__tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address cb date del email foreign gap graphic hi lb milestone name note num pb ptr ref time title unclear

derived-module-tei__tite: b cols i smcap sub sup ul

figures: figure formula

gaiji: g

linking: seg

transcr: handShift

Declaration

```
element time
{
   att.global.attributes,
   att.datable.attribute.calendar,
   att.datable.attribute.period,
   att.datable.w3c.attribute.when,
   att.typed.attributes,
   ( text | model.gLike | model.phrase | model.global )*
}
```

Example

```
    As he sat smiling, the
    quarter struck — <time when="11:45:00">the quarter to twelve</time>.
```

<ul> (underline) for capturing typographical feature: underlined glyphs.

Module derived-module-tei_tite

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

Used by model.hiLike

Contained by

core: abbr add addrLine author bibl date del desc editor email foreign head hi item l
label name note num p pubPlace publisher q ref resp speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener
salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list
listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element ul { att.global.attributes, macro.paraContent }
```

<unclear> contains a word, phrase, or passage which cannot be transcribed with
certainty because it is illegible or inaudible in the source.11.3.3.1. 3.4.3.

Module core

Attributes Attributes att.global (*@xml:id, @n, @xml:lang, @rend*)

@reason indicates why the material is hard to transcribe.

Status Optional

Datatype 1–∞ occurrences of data.word separated by whitespace

Values one or more words describing the difficulty, e.g. *faded,
background noise, passing truck, illegible, eccentric ductus.*

```
<div>
 <head>Rx</head>
 <p>500 mg <unclear
     reason="illegible">placebo</unclear>
 </p>
</div>
```

Used by model.choicePart model.pPart.transcriptional

Contained by

core: abbr add addrLine author bibl date del editor email foreign head hi item l label
name note num p pubPlace publisher q ref speaker stage time title unclear

derived-module-tei_tite: b i smcap sub sup ul

figures: cell

linking: ab seg

textstructure: byline closer dateline docAuthor docDate docEdition docImprint opener salute signed titlePart

May contain

core: abbr add address bibl cb cit date del desc email foreign gap graphic hi label lb list listBibl milestone name note num pb ptr q ref stage time title unclear

derived-module-tei_tite: b cols i ornament smcap sub sup ul

figures: figure formula table

gaiji: g

linking: seg

textstructure: floatingText

transcr: handShift

Declaration

```
element unclear
{
   attribute reason { list { data.word, data.word* } }?,
   att.global.attributes,
   macro.paraContent}
```

Example

```
and from time to time invited in like manner
his att<unclear>ention</unclear>
```

Here the last few letters of the word are hard to read.

Example

```
<u> ...and then <unclear reason="background-noise">Nathalie</unclear> said
... </u>
```

Note   The same element is used for all cases of uncertainty in the transcription of element content, whether for written or spoken material. For other aspects of certainty, uncertainty, and reliability of tagging and transcription, see chapter 21. . The <damage>, <gap>, <del>, <unclear> and <supplied> elements may be closely allied in use. See section 11.3.3.2. for discussion of which element is appropriate for which circumstance.

Schema tei_tite: unchanged components

add: (addition) contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.3.4.3.

addrLine: (address line) contains one line of a postal address.3.5.2. 2.2.4. 3.11.2.3.

address: contains a postal address, for example of a publisher, an organization, or an individual.3.5.2. 2.2.4. 3.11.2.3.

argument: A formal list or prose description of the topics addressed by a subdivision of a text.4.2. 4.6.

att.datable: provides attributes for normalization of elements that contain dates, times, or datable events.

att.declarable: provides attributes for those elements in the TEI Header which may be independently selected by means of the special purpose decls attribute.

att.global.facs: groups elements corresponding with all or part of an image, because they contain an alternative representation of it, typically but not necessarily a transcription of it.

att.sortable: provides attributes for elements in lists or groups that are sortable, but whose sorting key cannot be derived mechanically from the element content.

att.tableDecoration: provides attributes used to decorate rows or cells of a table.

att.transcriptional: provides attributes specific to elements encoding authorial or scribal intervention in a text when transcribing manuscript or similar sources.

author: in a bibliographic reference, contains the name(s) of an author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority. 3.11.2.2. 2.2.1.

back: (back matter) contains any appendixes, etc. following the main part of a text.4.7. 4.

bibl: (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged. 3.11.1. 2.2.7. 15.3.2.

body: (text body) contains the whole body of a single unitary text, excluding any front or back matter.4.

byline: contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.4.2.2. 4.5.

cb: (column break) marks the boundary between one column of a text and the next in a standard reference system.3.10.3.

cell: contains one cell of a table. 14.1.1.

closer: groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter.4.2.2. 4.2.

data.code: defines the range of attribute values expressing a coded value by means of a pointer to some other element which contains a definition for it.

data.count: defines the range of attribute values used for a non-negative integer value used as a count.

data.duration.w3c: defines the range of attribute values available for representation of a duration in time using W3C datatypes.

data.enumerated: defines the range of attribute values expressed as a single XML name taken from a list of documented possibilities.

data.language: defines the range of attribute values used to identify a particular combination of human language and writing system.

data.name: defines the range of attribute values expressed as an XML Name.

data.numeric: defines the range of attribute values used for numeric values.

data.pointer: defines the range of attribute values used to provide a single URI pointer to any other resource, either within the current document or elsewhere.

data.temporal.w3c: defines the range of attribute values expressing a temporal expression such as a date, a time, or a combination of them, that conform to the W3C *XML Schema Part 2: Datatypes* specification.

data.truthValue: defines the range of attribute values used to express a truth value.

data.word: defines the range of attribute values expressed as a single word or token.

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. 4.2.4. 4.2.2.

del: (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, annotator, or corrector.3.4.3.

desc: (description) contains a brief description of the object documented by its parent element, including its intended usage, purpose, or application where this is appropriate.22.4.4. 22.4.5. 22.4.6. 22.4.7.

div1: (level-1 text division) contains a first-level subdivision of the front, body, or back of a text.4.1.2.

div2: (level-2 text division) contains a second-level subdivision of the front, body, or back of a text.4.1.2.

div3: (level-3 text division) contains a third-level subdivision of the front, body, or back of a text.4.1.2.

div4: (level-4 text division) contains a fourth-level subdivision of the front, body, or back of a text.4.1.2.

div5: (level-5 text division) contains a fifth-level subdivision of the front, body, or back of a text.4.1.2.

div6: (level-6 text division) contains a sixth-level subdivision of the front, body, or back of a text.4.1.2.

div7: (level-7 text division) contains the smallest possible subdivision of the front, body or back of a text, larger than a paragraph.4.1.2.

docAuthor: (document author) contains the name of the author of the document, as given on the title page (often but not always contained in a byline).4.6.

docDate: (document date) contains the date of a document, as given (usually) on a title page.4.6.

docEdition: (document edition) contains an edition statement as presented on a title page of a document.4.6.

docImprint: (document imprint) contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.4.6.

docTitle: (document title) contains the title of a document, including all its constituents, as given on a title page.4.6.

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. 3.11.2.2.

email: (electronic mail address) contains an e-mail address identifying a location to which e-mail messages can be delivered. 3.5.2.

epigraph: contains a quotation, anonymous or attributed, appearing at the start or end of a section or on a title page. 4.2.3. 4.2. 4.6.

figure: groups elements representing or containing graphic information such as an illustration, formula, or figure.14.4.

floatingText: contains a single text of any kind, whether unitary or composite, which interrupts the text containing it at any point and after which the surrounding text resumes.4.3.2.

foreign: (foreign) identifies a word or phrase as belonging to some language other than that of the surrounding text. 3.3.2.1.

front: (front matter) contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.4.6. 4.

<**unclear**>

group: contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example the collected works of an author, a sequence of prose essays, etc.4. 4.3.1. 15.1.

handShift: marks the beginning of a sequence of text written in a new hand, or the beginning of a scribal stint.11.3.2.1.

head: (heading) contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc.4.2.1.

hi: (highlighted) marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made. 3.3.2.2. 3.3.2.

item: contains one component of a list.3.7. 2.5.

label: contains any label or heading used to identify part of a text, typically but not exclusively in a list or glossary.3.7.

lb: (line break) marks the start of a new (typographic) line in some edition or version of a text.3.10.3. 7.2.5.

lg: (line group) contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.3.12.1. 3.12. 7.2.5.

list: (list) contains any sequence of items organized as a list.3.7.

listBibl: (citation list) contains a list of bibliographic citations of any kind. 3.11.1. 2.2.7. 15.3.2.

macro.limitedContent: (paragraph content) defines the content of prose elements that are not used for transcription of extant materials.

macro.paraContent: (paragraph content) defines the content of paragraphs and similar elements.

macro.phraseSeq: (phrase sequence) defines a sequence of character data and phrase-level elements.

macro.phraseSeq.limited: (limited phrase sequence) defines a sequence of character data and those phrase-level elements that are not typically used for transcribing extant documents.

macro.specialPara: ('special' paragraph content) defines the content model of elements such as notes or list items, which either contain a series of component-level elements or else have the same structure as a paragraph, containing a series of phrase-level and inter-level elements.

milestone: marks a boundary point separating any kind of section of a text, typically but not necessarily indicating a point at which some part of a standard reference system changes, where the change is not represented by a structural element.3.10.3.

model.addrPart: groups elements such as names or postal codes which may appear as part of a postal address.

model.addressLike: groups elements used to represent a postal or e-mail address.

model.availabilityPart: groups elements such as licences and paragraphs of text which may appear as part of an availability statment

model.biblLike: groups elements containing a bibliographic description.

model.biblPart: groups elements which represent components of a bibliographic description.

model.choicePart: groups elements (other than <choice> itself) which can be used within a <choice> alternation.

model.common: groups common chunk- and inter-level elements.

model.dateLike: groups elements containing temporal expressions.

model.div1Like: groups top-level structural divisions.

model.div2Like: groups second-level structural divisions.

model.div3Like: groups third-level structural divisions.

model.div4Like: groups fourth-level structural divisions.

model.div5Like: groups fifth-level structural divisions.

model.div6Like: groups sixth-level structural divisions.

model.div7Like: groups seventh-level structural divisions.

model.divBottom: groups elements appearing at the end of a text division.

model.divBottomPart: groups elements which can occur only at the end of a text division.

model.divGenLike: groups elements used to represent a structural division which is generated rather than explicitly present in the source.

model.divLike: groups elements used to represent un-numbered generic structural divisions.

model.divPart: groups paragraph-level elements appearing directly within divisions.

model.divTop: groups elements appearing at the beginning of a text division.

model.divTopPart: groups elements which can occur only at the beginning of a text division.

model.divWrapper: groups elements which can appear at either top or bottom of a textual division.

model.egLike: groups elements containing examples or illustrations.

model.emphLike: groups phrase-level elements which are typographically distinct and to which a specific function can be attributed.

model.entryPart: groups elements appearing at any level within a dictionary entry.

model.frontPart: groups elements which appear at the level of divisions within front or back matter.

model.gLike: groups elements used to represent individual non-Unicode characters or glyphs.

model.global: groups elements which may appear at any point within a TEI text.

model.global.edit: groups globally available elements which perform a specifically editorial function.

model.glossLike: groups elements which provide an alternative name, explanation, or description for any markup construct.

model.graphicLike: groups elements containing images, formulae, and similar objects.

model.headLike: groups elements used to provide a title or heading at the start of a text division.

model.hiLike: groups phrase-level elements which are typographically distinct but to which no specific function can be attributed.

model.highlighted: groups phrase-level elements which are typographically distinct.

model.imprintPart: groups the bibliographic elements which occur inside imprints.

model.inter: groups elements which can appear either within or between paragraph-like elements.

model.lLike: groups elements representing metrical components such as verse lines.

model.labelLike: groups elements used to gloss or explain other parts of a document.

model.limitedPhrase: groups phrase-level elements excluding those elements primarily intended for transcription of existing sources.

model.linePart: groups transcriptional elements which appear within lines or zones of a source-oriented transcription within a <sourceDoc> element.

model.listLike: groups list-like elements.

model.measureLike: groups elements which denote a number, a quantity, a measurement, or similar piece of text that conveys some numerical meaning.

model.milestoneLike: groups milestone-style elements used to represent reference systems.

model.msItemPart: groups elements which can appear within a manuscript item description.

model.msQuoteLike: groups elements which represent passages such as titles quoted from a manuscript as a part of its description.

model.nameLike: groups elements which name or refer to a person, place, or organization.

model.nameLike.agent: groups elements which contain names of individuals or corporate bodies.

model.noteLike: groups globally-available note-like elements.

model.pLike: groups paragraph-like elements.

model.pLike.front: groups paragraph-like elements which can occur as direct constituents of front matter.

model.pPart.data: groups phrase-level elements containing names, dates, numbers, measures, and similar data.

model.pPart.edit: groups phrase-level elements for simple editorial correction and transcription.

model.pPart.editorial: groups phrase-level elements for simple editorial interventions that may be useful both in transcribing and in authoring.

model.pPart.transcriptional: groups phrase-level elements used for editorial transcription of pre-existing source materials.

model.personPart: groups elements which form part of the description of a person.

model.phrase: groups elements which can occur at the level of individual words or phrases.

model.ptrLike: groups elements used for purposes of location and reference.

model.publicationStmtPart: groups elements which may appear within the <publicationStmt> element of the TEI Header.

model.qLike: groups elements related to highlighting which can appear either within or between chunk-level elements.

model.quoteLike: groups elements used to directly contain quotations.

model.respLike: groups elements which are used to indicate intellectual or other significant responsibility, for example within a bibliographic element.

model.segLike: groups elements used for arbitrary segmentation.

model.stageLike: groups elements containing stage directions or similar things defined by the module for performance texts.

model.titlepagePart: groups elements which can occur as direct constituents of a title page, such as <docTitle>, <docAuthor>, <docImprint>, or <epigraph>.

name: (name, proper noun) contains a proper noun or noun phrase.3.5.1.

opener: groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter.4.2.4. 4.2.

p: (paragraph) marks paragraphs in prose.3.1. 7.2.5.

postscript: contains a postscript, e.g. to a letter.4.2.

pubPlace: (publication place) contains the name of the place where a bibliographic item was published.3.11.2.3.

publisher: provides the name of the organization responsible for the publication or distribution of a bibliographic item.3.11.2.3. 2.2.4.

q: (separated from the surrounding text with quotation marks) contains material which is marked as (ostensibly) being somehow different than the surrounding text, for any one of a variety of reasons including, but not limited to: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used.3.3.3.

resp: (responsibility) contains a phrase describing the nature of a person's intellectual responsibility, or an organization's role in the production or distribution of a work.3.11.2.2. 2.2.1. 2.2.2. 2.2.5.

respStmt: (statement of responsibility) supplies a statement of responsibility 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. May also be used to encode information about individuals or organizations which have played a role in the production or distribution of a bibliographic work.3.11.2.2. 2.2.1. 2.2.2. 2.2.5.

row: contains one row of a table. 14.1.1.

salute: (salutation) 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.4.2.4. 4.2.2.

seg: (arbitrary segment) represents any segmentation of text below the "chunk" level.16.3. 6.2. 7.2.5.

signed: (signature) contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text. 4.2.4. 4.2.2.

sp: (speech) An individual speech in a performance text, or a passage presented as such in a prose or verse text.3.12.2. 3.12. 7.2.2.

speaker: A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.3.12.2.

table: contains text displayed in tabular form, in rows and columns.14.1.1.

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.4. 15.1.

title: contains a title for any kind of work.3.11.2.2. 2.2.1. 2.2.5.

titlePage: (title page) contains the title page of a text, appearing within the front or back matter. 4.6.

titlePart: contains a subsection or division of the title of a work, as indicated on a title page.4.6.

Acknowledgments

The TEI Tite is simply a synthesis of work done at the

- University of Michigan Digital Library Production Service,

- University of Virginia Digital Library Production Service,

- and the California Digital Library

and represented in their documents

- *Minimum standards for text capture,*

- *Text Encoding Guidelines for Keyboarding Vendors*,

- and *CDL TEI Base Encoding Guidelines*,

respectively. Many thanks to the institutions and individuals responsible for sharing their experience and expertise for the benefit of the TEI community at large.

Also, thank you to members of the TEI Special Interest Group on Libraries who provided very valuable corrections and suggestions.