TEI Tite — A recommendation for off-site text encoding

Perry Trolard, for TEI Consortium

Version 1.0 — May 2009

1 Introduction

This document specifies how TEI Tite should be applied. 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-encoded documents are TEI documents, and TEI Tite, with the exception of convenience elements (<b>, <i>, <ul>, <sup>, <sub>, <smcap>, <cols> and <ornament>, all of which can be converted back to canonical TEI), is a pure subset of the TEI. That is, it 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) 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 element, but as a <head> within the appropriate . 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 . As a rough rule of thumb, if there is a title accompanying the division, use the 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 elements with type attributes of scene or act, respectively. Cast lists can likewise be encoded using and `type="castlist"`.

Prologues and epilogues can be treated as <sp>s of their own, unless their structure would be better represented by nested 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 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 *TEI Text Encoding in Libraries* Guidelines

As of May 2009, the TEI Special Interest Group (SIG) on Libraries is in the process of revising its *TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices* (GBP) document, which creates common definitions of "levels of encoding" based on depth of markup applied. A draft of this document is available at the Libraries SIG's wiki. 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 GBP levels, TEI Tite would sit between three and four: it requires use of all the elements from level three plus additional ones, but requires fewer elements than level four. Relative to level three, "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.

Because Tite is closer to it, it's level four ("Basic Content Analysis") that provides the most useful comparison. The folowing items represent instances where Tite is *less* ambitious than level four:

- 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

As of this point, it seems that bringing Tite-encoded documents up to GBP level four would simply require additional application of markup, not significant reworking of markup, and in that way Tite seems compatible with the GBP.

Not mentioned above is one key but purposeful incompatibility: Tite's lack of a header. A TEI header must be added, and the root <TEI> element used for compatibility with the GBP, and with the TEI abstract model in general.

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.

Formal specification

Schema tei_tite: changed components

---

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

Module tei

Members bibl listBibl

Attributes Global attributes only

---

att.editLike provides attributes describing the nature of a encoded scholarly intervention or interpretation of any kind.

Module tei

Members date gap time unclear

Attributes att.dimensions (*@unit*, *@quantity*, *@extent*, *@precision*, *@scope*) (att.ranging (*@atLeast*, *@atMost*, *@min*, *@max*)) att.responsibility (*@cert*, *@resp*)

@source contains a list of one or more pointers indicating sources supporting the given intervention or interpretation.

Status Mandatory when applicable

Datatype 1–∞ occurrences of `xsd:anyURI` separated by whitespace

Values A space-delimited series of sigla; each sigil should correspond to a witness or witness group and occur as the value of the xml:id attribute on a <witness> or <msDesc> element elsewhere in the document.

---

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

Module tei

Members b cols i ornament smcap sub sup ul

Attributes In addition to global 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 CORS.

`@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 `token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }` separated by whitespace

Values the value 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 `xsd: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 If no value is specified for xml:lang, the xml:lang value for the immediately enclosing element is inherited; for this reason, a value should always be specified on the outermost element (<TEI>).

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

Status Optional

Datatype 1–∞ occurrences of `token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }` 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.

@xml:base provides a base URI reference with which applications can resolve relative URI references into absolute URI references.
Status Optional
Datatype xsd:anyURI
Values any syntactically valid URI reference.

```
<div
   type="bibl">
  <head>Bibliography</head>
  <listBibl
     xml:base="http://www.lib.ucdavis.edu/BWRP/Works/">
    <bibl
       n="1">
      <author>
        <name>Landon, Letitia Elizabeth</name>
      </author>
      <ref
         target="LandLVowOf.sgm">
        <title>The Vow of the Peacock</title>
      </ref>
    </bibl>
    <bibl
       n="2">
      <author>
        <name>Compton, Margaret Clephane</name>
      </author>
      <ref
         target="NortMIrene.sgm">
        <title>Irene, a Poem in Six Cantos</title>
      </ref>
    </bibl>
    <bibl
       n="3">
      <author>
        <name>Taylor, Jane</name>
      </author>
      <ref
         target="TaylJEssay.sgm">
        <title>Essays in Rhyme on Morals and Manners</title>
      </ref>
    </bibl>
  </listBibl>
</div>
```

@xml:space signals an intention about how white space should be managed by applications.
Status Optional
Legal values are: **default** the processor should treat white space according to the default XML white space handling rules
preserve the processor should preserve unchanged any and all white space in the source
Note The XML specification provides further guidance on the use of this attribute.

---

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

Module tei

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

Attributes In addition to global attributes

@type characterizes the element in some sense, using any convenient classification scheme or typology.
Status Optional
Datatype `xsd:Name`

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

Module derived-module-tei_tite

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.hiLike

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

Declaration

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

<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

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

@ed indicates the edition or version in which the change in columnar layout is located at this point
Status Optional
Datatype `xsd:anyURI`

Used by model.milestoneLike

May contain Empty element

Declaration

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

<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.

Module core

In addition to global attributes att.editLike (*@source*) (att.dimensions (*@unit*, *@quantity*, *@extent*, *@precision*, *@scope*) (att.ranging (*@atLeast*, *@atMost*, *@min*, *@max*))) (att.responsibility (*@cert*, *@resp*))

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

Status Optional

Datatype 1–∞ occurrences of `token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }` separated by whitespace

Values any short indication of the reason for the omission.

Used by model.global.edit

May contain

core: desc

Declaration

```
element gap
{
   attribute reason
   {
      list
      {
         token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" },
         token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }*
      }
   }?,
   att.global.attributes,
   att.editLike.attribute.source,
   att.dimensions.attributes,
   model.glossLike*
}
```

Example

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

Example

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

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 PHCOMB for discussion of which element is appropriate for which circumstance.

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

Module derived-module-tei_tite

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.hiLike

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

Declaration

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

<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 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

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.inter model.titlepagePart

May contain Character data only

Declaration

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

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

Module derived-module-tei_tite

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.hiLike

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

Declaration

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

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

Module derived-module-tei_tite

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.hiLike

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

Declaration

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

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

Module derived-module-tei_tite

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.hiLike

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

Declaration

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

 (underline) for capturing typographical feature: underlined glyphs.

Module derived-module-tei_tite

In addition to global attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:base*, *@xml:space*)

Used by model.hiLike

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

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.

Module core

In addition to global attributes att.editLike (*@source*) (att.dimensions (*@unit*, *@quantity*, *@extent*, *@precision*, *@scope*) (att.ranging (*@atLeast*, *@atMost*, *@min*, *@max*))) (att.responsibility (*@cert*, *@resp*))

@reason indicates why the material is hard to transcribe.

Status Optional

Datatype 1–∞ occurrences of `token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }` 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.pPart.transcriptional model.choicePart

May contain

core: abbr address bibl cb cit date desc email foreign gap graphic hi label lb list listBibl measureGrp 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

Declaration

```
element unclear
{
   attribute reason
   {
      list
      {
         token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" },
         token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }*
      }
   }?,
   att.global.attributes,
```

```
   att.editLike.attribute.source,
   att.dimensions.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 CE. The <damage>, <gap>, <del>, <unclear> and <supplied> elements may be closely allied in use. See section PHCOMB for discussion of which element is appropriate for which circumstance.

Schema tei_tite: unchanged 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.
abbr: (abbreviation) contains an abbreviation of any sort.
addrLine: (address line) contains one line of a postal address.
address: contains a postal address, for example of a publisher, an organization, or an individual.
argument: A formal list or prose description of the topics addressed by a subdivision of a text.
att.ascribed: provides attributes for elements representing speech or action that can be ascribed to a specific individual.
att.canonical: provides attributes which can be used to associate a representation such as a name or title with canonical information about the object being named or referenced.
att.datable: provides attributes for normalization of elements that contain dates, times, or datable events.
att.datable.w3c: provides attributes for normalization of elements that contain datable events using the W3C datatypes.
att.declaring: provides attributes for elements which may be independently associated with a particular declarable element within the header, thus overriding the inherited default for that element.
att.dimensions: provides attributes for describing the size of physical objects.
att.internetMedia: provides attributes for specifying the type of a computer resource using a standard taxonomy.
att.measurement: provides attributes to represent a regularized or normalized measurement.
att.naming: provides attributes common to elements which refer to named persons, places, organizations etc.
att.placement: provides attributes for describing where on the source page or object a textual element appears.

att.pointing: defines a set of attributes used by all elements which point to other elements by means of one or more URI references.
att.ranging: provides attributes for describing numerical ranges.
att.responsibility: provides attributes indicating who is responsible for something asserted by the markup and the degree of certainty associated with it.
att.sourced: provides attributes identifying the source edition from which some encoded feature derives.
att.spanning: provides attributes for elements which delimit a span of text by pointing mechanisms rather than by enclosing it.
att.tableDecoration: provides attributes used to decorate rows or cells of a table.
att.translatable: provides attributes used to indicate the status of a translatable portion of an ODD document.
author: in a bibliographic reference, contains the name(s) of the author(s), personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.
back: (back matter) contains any appendixes, etc. following the main part of a text.
bibl: (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.
biblScope: (scope of citation) defines the scope of a bibliographic reference, for example as a list of page numbers, or a named subdivision of a larger work.
body: (text body) contains the whole body of a single unitary text, excluding any front or back matter.
byline: contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.
cb: (column break) marks the boundary between one column of a text and the next in a standard reference system.
cell: contains one cell of a table.
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.
closer: groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter.
date: contains a date in any format.
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.
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.
div1: (level-1 text division) contains a first-level subdivision of the front, body, or back of a text.
div2: (level-2 text division) contains a second-level subdivision of the front, body, or back of a text.
div3: (level-3 text division) contains a third-level subdivision of the front, body, or back of a text.
div4: (level-4 text division) contains a fourth-level subdivision of the front, body, or back of a text.

div5: (level-5 text division) contains a fifth-level subdivision of the front, body, or back of a text.
div6: (level-6 text division) contains a sixth-level subdivision of the front, body, or back of a text.
div7: (level-7 text division) contains the smallest possible subdivision of the front, body or back of a text, larger than a paragraph.
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).
docDate: (document date) contains the date of a document, as given (usually) on a title page.
docEdition: (document edition) contains an edition statement as presented on a title page of a document.
docImprint: (document imprint) contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.
docTitle: (document title) contains the title of a document, including all its constituents, as given on a title page.
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.
email: (electronic mail address) contains an e-mail address identifying a location to which e-mail messages can be delivered.
epigraph: contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.
figure: groups elements representing or containing graphic information such as an illustration or figure.
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.
foreign: (foreign) identifies a word or phrase as belonging to some language other than that of the surrounding text.
formula: contains a mathematical or other formula.
front: (front matter) contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.
g: (character or glyph) represents a non-standard character or glyph.
graphic: indicates the location of an inline graphic, illustration, or figure.
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.
head: (heading) contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc.
hi: (highlighted) marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.
item: contains one component of a list.
l: (verse line) contains a single, possibly incomplete, line of verse.
label: contains the label associated with an item in a list; in glossaries, marks the term being defined.
lb: (line break) marks the start of a new (typographic) line in some edition or version of a text.
lg: (line group) contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.

list: (list) contains any sequence of items organized as a list.
listBibl: (citation list) contains a list of bibliographic citations of any kind.
macro.anyXML: defines a content model within which any XML elements are permitted
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.
measureGrp: (measure group) contains a group of dimensional specifications which relate to the same object, for example the height and width of a manuscript page.
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.
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.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.entryPart.top: groups high level elements within a structured 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.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.
note: contains a note or annotation.
num: (number) contains a number, written in any form.
opener: groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter.
p: (paragraph) marks paragraphs in prose.
pb: (page break) marks the boundary between one page of a text and the next in a standard reference system.
postscript: contains a postscript, e.g. to a letter.
ptr: (pointer) defines a pointer to another location.
pubPlace: (publication place) 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.
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.

ref: (reference) defines a reference to another location, possibly modified by additional text or comment.
relatedItem: contains or references some other bibliographic item which is related to the present one in some specified manner, for example as a constituent or alternative version of it.
resp: (responsibility) contains a phrase describing the nature of a person's intellectual responsibility.
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.
row: contains one row of a table.
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.
seg: (arbitrary segment) represents any segmentation of text below the "chunk" level.
signed: (signature) contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text.
sp: (speech) An individual speech in a performance text, or a passage presented as such in a prose or verse text.
speaker: A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.
stage: (stage direction) contains any kind of stage direction within a dramatic text or fragment.
table: contains text displayed in tabular form, in rows and columns.
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.
time: contains a phrase defining a time of day in any format.
title: contains a title for any kind of work.
titlePage: (title page) contains the title page of a text, appearing within the front or back matter.
titlePart: contains a subsection or division of the title of a work, as indicated on a title page.
trailer: contains a closing title or footer appearing at the end of a division of a text.