



Apex CoVantage, LLC
200 Presidents Plaza
198 Van Buren Street
Herndon, Virginia 20170-5338
USA

Content Solutions



Apex TEI -TITE Guidelines For TEI Access Project



Table of Contents

1.0	General guidelines	1-5
1.1	Introduction.....	1-5
1.2	General Requirements	1-6
1.3	Global Text Structure.....	1-7
1.4	Types of Text and Defaults.....	1-13
1.5	Block-level Features	1-16
1.6	Phrase-level Features	1-19
2.0	Appendices.....	2-22
2.1	Appendix B Acknowledgments.....	2-22
3.0	Element Descriptions	3-23
3.1	<ab>	3-23
3.2	<abbr>	3-23
3.3	<add>	3-23
3.4	<addrLine>	3-23
3.5	<address>.....	3-23
3.6	<argument>	3-23
3.7	<author>	3-23
3.8		3-24
3.9	<back>.....	3-24
3.10	<bibl>.....	3-24
3.11	<body>.....	3-24
3.12	<byline>	3-24
3.13	<cb>.....	3-24
3.14	<cell>	3-24
3.15	<cit>.....	3-24
3.16	<closer>.....	3-25
3.17	<cols>	3-25
3.18	<date>	3-25
3.19	<dateline>.....	3-25
3.20	3-25
3.21	<desc>.....	3-25
3.22	<div1>.....	3-26
3.23	<div2>.....	3-26
3.24	<div3>.....	3-26
3.25	<div4>.....	3-26
3.26	<div5>.....	3-26
3.27	<div6>.....	3-26
3.28	<div7>.....	3-26
3.29	<docAuthor>	3-27
3.30	<docDate>	3-27
3.31	<docEdition>.....	3-27
3.32	<docImprint>.....	3-27
3.33	<docTitle>.....	3-27
3.34	<editor>	3-27
3.35	<email>.....	3-27
3.36	<epigraph>	3-28

3.37	<figure>	3-28
3.38	</figure><floatingText>	3-28
3.39	<formula>	3-28
3.40	<front>	3-28
3.41	<g>	3-28
3.42	<gap>	3-28
3.43	<graphic>	3-29
3.44	<group>	3-29
3.45	<handShift>	3-29
3.46	<head>	3-29
3.47	<hi>	3-29
3.48	<i>	3-29
3.49	<item>	3-29
3.50	<l>	3-29
3.51	label>	3-30
3.52	<lb>	3-30
3.53	<lg>	3-30
3.54	<list>	3-30
3.55	<listBibl>	3-30
3.56	<milestone>	3-30
3.57	<name>	3-30
3.58	<note>	3-30
3.59	<num>	3-31
3.60	<opener>	3-31
3.61	<ornament>	3-31
3.62	<p>	3-31
3.63	<pb>	3-31
3.64	<postscript>	3-31
3.65	<ptr>	3-31
3.66	<pubPlace>	3-32
3.67	<publisher>	3-32
3.68	<q>	3-32
3.69	<ref>	3-32
3.70	<relatedItem>	3-32
3.71	<resp>	3-32
3.72	<respStmt>	3-32
3.73	<row>	3-33
3.74	<salute>	3-33
3.75	<seg>	3-33
3.76	<signed>	3-33
3.77	<smcap>	3-33
3.78	<sp>	3-33
3.79	<speaker>	3-33
3.80	<stage>	3-34
3.81	<sub>	3-34
3.82	<sup>	3-34
3.83	<table>	3-34
3.84	<text>	3-34
3.85	<time>	3-34
3.86	<title>	3-34

3.87	<titlePage>	3-34
3.88	<titlePart>	3-35
3.89	<trailer>	3-35
3.90	3-35
3.91	<unclear>.....	3-35
4.0	Exhibits	4-36
4.1	Exhibit 1	4-36
4.2	Exhibit 2	4-38
4.3	Exhibit 3	4-39
4.4	Exhibit 4	4-40
4.5	Exhibit 5	4-41
4.6	Exhibit 6	4-42
4.7	Exhibit 7	4-43

1.0 General guidelines

1.1 Introduction

This document specifies how TEI TITE should be applied in the case of AccessTEI projects. It outlines the basic structures and elements that should be used and contains suggestions that will help standardize tagging procedure and thus maintain the efficiency of workflows.

This document is organized in roughly the structure of a TEI document itself. 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), then 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 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 (, <i>, , <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 official TEI Guidelines. TITE diverges from general TEI because it seeks uniformity of encoding through constraint, via its stripped-down tag set and via this specification, whereas the TEI in general is famously tolerant of multiple methods of encoding a given feature.

A general idea of the TEI class system will aid in understanding the tagging process. Elements within TEI and TITE are usually formally classified in some way. Classes are used to express two distinct kinds of commonality among elements. The elements of a class may share some set of attributes, or they may appear in the same locations in a content model. A class is known as an attribute class if its members share attributes, and as a model class if its members appear in the same locations. In either case, an element is said to inherit properties from any classes of which it is a member.

Classes (and therefore elements which are members of those classes) may also inherit properties from other classes. For example, supposing that class A is a member (or a subclass) of class B, any element which is a member of class A will inherit not only the properties defined by class A, but also those defined by class B. In such a situation, we also say that class B is a superclass of class A. The properties of a superclass are inherited by all members of its subclasses.

The sample material types discussed in this document include letters, verse, drama, newspapers, and directories, but TITE can be used to encode anything that 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.

1.2 General Requirements

1.2.1 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.

1.2.2 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.

1.2.3 Character Encoding

Characters should be encoded in UTF-8. This is the preferred method as it is able to represent any character in the Unicode standard, yet is backwards compatible with ASCII. 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).

1.2.4 Accuracy and Verification

The standard for accuracy of transcription should be at least 99.95%.

1.2.5 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, coders should decide on what solution best fits the situation, implement the solution and document what is done. To this end, each encoded file should be accompanied by a document with notes that reference features of a document for which there were no clear encoding method and how these features were handled by encoders.

1.3 Global Text Structure

1.3.1 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>
    <!-- front matter -->
  </front>
  <body>
    <!-- body of text -->
  </body>
  <back>
    <!-- back matter -->
  </back>
</text>
```

The [text](#)'s xml:id attribute should contain a unique identifier for the document being encoded. Customers may supply their own specifications for unique identifiers if they are valid, but the default formula for xml:id will be:

```
<text xml:id="text#.batch#.date">
```

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.

1.3.2 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 nests within the [text](#) element and contains 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 a 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>
  <!-- front matter for the group -->
  </front>
  <group>
    <text>
      • <front>
      • <!-- front matter of first text -->
      • </front>
      • <body>
      • <!-- body of first text -->
      • </body>
      • <back>
      • <!-- back matter of first text -->
      • </back>
    • </text>
    • <text>
      • <front>
      • <!-- front matter of second text -->
    >
      • </front>
      • <body>
      • <!-- body of second text -->
      • </body>
      • <back>
      • <!-- back matter of second text -->
      • </back>
    • </text>
    • <!-- more texts or groups of texts here -->
  • </group>
  • <back>
  <!-- back matter for the 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) `<divn>` with an appropriate type attribute.

***Note:** `<divn>` refers to a division of an undetermined number, used from this point on in the document to refer to numbered division tags in the general sense.

1.3.3 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. For each job, a constrained enumerated list of type values will have to be determined, either by instruction from the client or Apex's understanding of what the client might need or a combination of the two. However, the following are some general examples of appropriate division types:

- 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)
- running

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:

```
<div n="III" type="part">
  <head>III: It Awakes</head>
  <!-- ... -->
</div>
```

1.3.3.1 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 `<divn>` element, but as a [head](#) within the appropriate `<divn>`. For half-title pages and similar flytitle pages, see the following section 1.3.4 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.

1.3.4 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—often copyright information—should be included as well (after a [pb](#)).

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. Encode half-titles 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).

1.4 Types of Text and Defaults

TITE is equipped to support basic encoding of several types of text. TITE has special elements for letters, verse, drama, and manuscripts; but it can support a wide range of formats like books, newspapers, directories, and pamphlets in addition to other similar printed material. The following sections present explanations and samples of default tagging for some basic formats that may be encountered in AccessTEI projects.

1.4.1 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

1.4.2 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 the type of line is known, use the corresponding 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 `<divn>`. As a rough rule of thumb, if there is a title accompanying the division, use the `<divn>` 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 line breaks motivated by page layout. Line breaks motivated by layout 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. " </1>  
</1g>
```

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.

1.4.3 Newspapers

Tite includes the elements `cols` and `cb` which are well suited for the multicolumn 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.

Refer to example [Exhibit_Newspaper.xml](#).

1.4.4 Books

Refer to example [Exhibit_Book.xml](#).

1.4.5 Journal or Serial

Refer to examples [Exhibit_1_Journal.xml](#) and [Exhibit_2_Journal.xml](#).

1.4.6 Magazine

Refer to example [Exhibit_Magazine.xml](#).

1.4.7 Transcripts

Refer to example [Exhibit_Transcripts.xml](#).

1.4.8 Monograph

Refer to example [Exhibit_Monograph.xml](#).

1.4.9 Directories

Individual listings in a directory, such as the name entries within a city directory as in the example below, will be treated as `item` elements within an item list.

Refer to example [Directory.xml](#)

1.4.10 Dictionary

Refer to example [Exhibit_Dictionaries.xml](#).

1.4.11 Encyclopedia

Refer to example [Exhibit_Encyclopedias.xml](#).

1.4.12 Manuscripts

TITE allows for encoding of handwritten material. In transcribing, use the element `<handShift>` to mark the beginning of a sequence of text written in a new hand, or the beginning of a section of handwriting. Also, use the elements `<add>` and `` for anything added or crossed out by hand. In the case of text that has been crossed out with new text added, key the deleted text first and then follow with the added text.

Refer to example [Exhibit_Manuscript.xml](#).

1.5 Block-level Features

1.5.1 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>
    <!-- ... -->
  </p>
  <q>
    <floatingText>
      <body>
        <lg type="poem">
<!-- poem -->
          </lg>
        </body>
      </floatingText>
    </q>
  <p>
    <!-- ... -->
  </p>
</div1>

```

If present, transcribe all quotation marks or other delimiters inside the `q` element. For guidelines concerning quotation marks within the `q` element, refer to Section 1.6.2.

1.5.2 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.

```

<div3 type="section" n="2">
  <p>
    <!--...-->
    <ref target="fig1">Fig. I</ref>
    <!--...-->
  </p>
  <figure xml:id="fig1">
    <graphic url="images/fig1.png"/>
    <ab type="caption"><smcap>FIG.</smcap>
    I.
    <!--...-->
  </ab>
  </figure>
</div3>

```

1.5.3 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.

1.5.4 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 (as is often the case for marginal notes), supply a [ptr](#) element in a reasonable place in the text running beside the note.

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.

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)

1.5.5 '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>

```

1.5.6 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.)

See use of `ab` in caption example for figure references.

1.6 Phrase-level Features

1.6.1 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 with a `rend` attribute.

1.6.2 Phrase-level Quotation

For passages set off by quotation marks or another delimiter, use the **q** element. The delimiter inside the tag should use the special UTF-8 characters 0201C for opening tags and 0201D for closing tags, unless the client prefers otherwise.

1.6.3 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.

1.6.4 Uncertain Segments

The **seg** element is the phrase-level analogue to the **ab** element. If a phraselevel 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).

1.6.5 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.

1.6.6 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.' `pb` should be considered standard for page breaks, except for in the case of poetry, math, or other non-prose cases. In this case, line breaks should be captured using the `lb` element.

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.

2.0 Appendices

2.1 Appendix B Acknowledgments

The TEI TITE described in this document is based upon the masterwork of Perry Trolard 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.

Apex has modified this work in order to create a set of production guidelines to clarify and inform the work that will be done in a series of AccessTEI projects, a collaborative effort between Apex and the TEI Consortium.

3.0 Element Descriptions

3.1 <ab>

<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.

3.2 <abbr>

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

3.3 <add>

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

3.4 <addrLine>

<addrLine> (address line) contains one line of a postal address. Please refer Exhibit 1 for coding example.

3.5 <address>

<address> contains a postal address, for example of a publisher, an organization, or an individual. Please refer Exhibit 1 for coding example.

3.6 <argument>

<argument> A formal list or prose description of the topics addressed by a subdivision of a text. Please refer Exhibit 1 for coding example.

3.7 <author>

<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. Please refer Exhibit 1 for coding example.

3.8

 (bold) for capturing typographical feature: bold glyphs.

3.9 <back>

<back> (back matter) contains any appendixes, etc. following the main part of a text.

3.10 <bibl>

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

3.11 <body>

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

3.12 <byline>

<byline> contains the primary statement of responsibility given for a work on its title page or at the head or end of the work. Please refer Exhibit 2 for coding example.

3.13 <cb>

<cb> (column break) marks the boundary between one column of a text and the next in a standard reference system. Please refer Exhibit 2 for coding example.

3.14 <cell>

<cell> contains one cell of a table.

3.15 <cit>

<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.

3.16 <closer>

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

3.17 <cols>

<cols> (columns) with the 'n' attribute (denoting new number of columns) is used to mark where a document changes columnar layout. Please refer Exhibit 1 for coding example.

3.18 <date>

<date> contains a date in any format.

3.19 <dateline>

<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.

3.20

 (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.

3.21 <desc>

<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.

3.22 <div1>

<div1> (level-1 text division) contains a first-level subdivision of the front, body, or back of a text. Please refer Exhibit 3 for coding example.

3.23 <div2>

<div2> (level-2 text division) contains a second-level subdivision of the front, body, or back of a text. Please refer Exhibit 3 for coding example.

3.24 <div3>

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

3.25 <div4>

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

3.26 <div5>

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

3.27 <div6>

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

3.28 <div7>

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

3.29 <docAuthor>

<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).

3.30 <docDate>

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

3.31 <docEdition>

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

3.32 <docImprint>

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

3.33 <docTitle>

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

3.34 <editor>

<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.

3.35 <email>

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

3.36 <epigraph>

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

3.37 <figure>

<figure> groups elements representing or containing graphic information such as an illustration or figure. Please refer Exhibit 4 for coding example.

3.38 </figure><floatingText>

<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.

3.39 <formula>

<formula> contains a mathematical or other formula.

3.40 <front>

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

3.41 <g>

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

3.42 <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.

3.43 <graphic>

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

3.44 <group>

<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.

3.45 <handShift>

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

3.46 <head>

<head> (heading) contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc. Please refer Exhibit 5 for coding example.

3.47 <hi>

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

3.48 <i>

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

3.49 <item>

<item> contains one component of a list.

3.50 <l>

<l> (verse line) contains a single, possibly incomplete, line of verse. Please refer Exhibit 5 for coding example.

3.51 <label>

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

3.52 <lb>

<lb> (line break) marks the start of a new (typographic) line in some edition or version of a text.

3.53 <lg>

<lg> (line group) contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc. Please refer Exhibit 5 for coding example.

3.54 <list>

<list> (list) contains any sequence of items organized as a list.

3.55 <listBibl>

<listBibl> (citation list) contains a list of bibliographic citations of any kind.

3.56 <milestone>

<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.

3.57 <name>

<name> (name, proper noun) contains a proper noun or noun phrase.

3.58 <note>

<note> contains a note or annotation.

3.59 <num>

<num> (number) contains a number, written in any form. Please refer Exhibit 3 for coding example.

3.60 <opener>

<opener> groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter. Please refer Exhibit 6 for coding example.

3.61 <ornament>

<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 empty; if the ornament can be represented with characters, include these in the element.

3.62 <p>

<p> (paragraph) marks paragraphs in prose.

3.63 <pb>

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

3.64 <postscript>

<postscript> contains a postscript, e.g. to a letter.

3.65 <ptr>

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

3.66 <pubPlace>

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

3.67 <publisher>

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

3.68 <q>

<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.

3.69 <ref>

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

3.70 <relatedItem>

<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.

3.71 <resp>

<resp> (responsibility) contains a phrase describing the nature of a person's intellectual responsibility.

3.72 <respStmt>

<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.

3.73 <row>

<row> contains one row of a table.

3.74 <salute>

<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. Please refer Exhibit 6 for coding example.

3.75 <seg>

<seg> (arbitrary segment) represents any segmentation of text below the 'chunk' level.

3.76 <signed>

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

3.77 <smcap>

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

3.78 <sp>

<sp> (speech) An individual speech in a performance text, or a passage presented as such in a prose or verse text. Please refer Exhibit 7 for coding example.

3.79 <speaker>

<speaker> A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment. Please refer Exhibit 7 for coding example.

3.80 <stage>

<stage> (stage direction) contains any kind of stage direction within a dramatic text or fragment. Please refer Exhibit 7 for coding example.

3.81 <sub>

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

3.82 <sup>

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

3.83 <table>

<table> contains text displayed in tabular form, in rows and columns.

3.84 <text>

<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.

3.85 <time>

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

3.86 <title>

<title> contains a title for any kind of work.

3.87 <titlePage>

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

3.88 <titlePart>

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

3.89 <trailer>

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

3.90

 (underline) for capturing typographical feature: underlined glyphs.

3.91 <unclear>

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

4.0 Exhibits

4.1 Exhibit 1

ADMINISTRATIVE HANDLING OF THE NARCOTIC ADDICT: ITS BENEFITS AND DANGERS.

ERNEST S. BISHOP, M. D., F. A. C. P.,

Clinical Professor of Medicine, New York Polyclinic Medical School, New York City.

Read before Food and Drugs Section, American Public Health Association, at New Orleans, La., October 27, 1919.

Dr. Bishop asserts again that narcotic drug addiction is a disease; the laws make it a crime. The appalling extent of illicit drug traffic is due largely to this mistake. Forcible control is a fundamental error. The existing laws make it hard for the physician to know where he stands in his treatment. Here is a strong plea for sensible reform.

A VERY serious question is arising from the narcotic drug situation. It constitutes an important problem for immediate consideration. This simple. Narcotic addiction, supposedly a "habit" or vice or indulgence, as it was then generally considered, seemed to require for its suppression merely forcible

<div1 type="article">

<head type="main">ADMINISTRATIVE HANDLING OF THE
NARCOTIC ADDICT: ITS BENEFITS AND DANGERS.</head>

<author><smcap>Ernest S. Bishop</smcap>, M. D., F. A. C.
P.,</author>

<address>

<addrLine><i>Clinical Professor of Medicine,</i></addrLine>

<addrLine><i>New York Polyclinic Medical School,</i></addrLine>

<addrLine><i>New York City.</i></addrLine></address>

<argument><p>Read before Food and Drugs Section, American
Public Health Association, at New Orleans, La., October 27,
1919.</p>

<p>Dr. Bishop asserts again that narcotic drug addiction is a disease;
the laws make it a crime. The appalling extent of illicit drug traffic is
due largely to this mistake. Forcible control is a fundamental error.
The existing laws make it hard for the physician to know where he

stands in his treatment. Here is a strong plea for sensible reform.</p></argument>

<cols n="2">

<p>A VERY serious question is arising from the narcotic drug situation. It constitutes an important problem for immediate ...consideration. The <cb/>simple. Narcotic addition, supposedly a “habitÉD; or vice or indulgence, as it ...</p>

4.2 Exhibit 2

The Pressure of Light

A NEW SCIENTIFIC DISCOVERY AND ITS WONDERFUL CONSEQUENCES

BY WALDEMAR KAEMPFERT

WITH the aid of instruments that feel what our hands can never feel, and see what our eyes can never see, the modern physicist has criticized the zodiacal light, that shimmering fabric which is mysteriously spread on the western horizon during the clear evenings of winter and spring? What is the message

```
<div1 type="article">
<head type="main">The Pressure of Light</head>
<head type="sub"><smcap>A NEW SCIENTIFIC DISCOVERY
AND ITS WONDERFUL CONSEQUENCES</smcap></head>
<byline><i>By <author>WALDEMAR
KAEMPFERT</author></byline>
```

4.3 Exhibit 3

ACT I.

SCENE I. *Venice. A street.**Enter* ANTONIO, SALARINO, and SALANIO.

Ant. In sooth, I know not why I am so sad :
 It wearies me ; you say it wearies you ;
 But how I caught it, found it, or came by it,
 What stuff 't is made of, whereof it is born,
 I am to learn ;
 And such a want-wit sadness makes of me,
 That I have much ado to know myself.

5

Salar. Your mind is tossing on the ocean ;
 There, where your argosies with portly sail,

B

1

```

<div1 type="act" n="1">
<head type="main">ACT I.</head>
<div2 type="scene" n="1">
<head type="main"><smcap>SCENE</smcap> I. <i>Venice. A
street.</i></head>
<stage><i>Enter</i> <smcap>ANTONIO, SALARINO</smcap>,
<i>and</i> <smcap>SALANIO</smcap>.</stage>
<sp><speaker><i>Ant.</i></speaker>
<lg><l>In sooth, I know not why I am so sad :</l>
<l>It wearies me ; you say it wearies you ;</l>
<l>But how I caught it, found it, or came by it,</l>
<l>What stuff 't is made of, whereof it is born,</l>
<l>I am to learn ; 5 <num n="5" rend="right">5</num></l>
<l>And such a want-wit sadness makes of me,</l>
<l>That I have much ado to know myself.</l></lg></sp>
<sp><speaker><i>Salar.</i></speaker>
<lg><l>Your mind is tossing on the ocean ;</l>
<l>There, where your argosies with portly sail,</l></lg>
</sp>
<ab type="printer's_number">B</ab>
</div2>
</div1>

```

4.4 Exhibit 4



FIG. 1.—Head of Siamese Deer (*Cervus schomburghii*), showing antlers.

<figure xml:id="fig1">
<graphic url="images/encyclopedia_fig1.png"/>
<ab type="caption"><smcap>FIG.</smcap> I.—Head of
Siamese Deer (<i>Cervus schomburgkis</i>), showing antlers.</ab>
</figure>

4.5 Exhibit 5

TWILIGHT

SHE

*T*HERE is no mystery, she saith.
Look down into the pits of death,
Where blue sleep lieth, fold on fold;
Look up, where God's white pardon waits
Forever at the open gates;
See, then, my hair all gold,
Good for a man to kiss and hold
And play with, as the hours spin by. . . .

```
<div1 type="poetry">  
<head type="main">TWILIGHT</head>  
<lg type="verse">  
<head type="main">SHE</head>  
<l><i>THERE is no mystery, she saith.</i></l>  
<l><i>Look down into the pits of death,</i></l>  
....  
<l><i>And play with, as the hours spin by ....</i></l>  
</lg>
```

4.6 Exhibit 6

MESSAGE.

To the Senate and House of Representatives:

I submit for the information of Congress the report of the commission appointed by me to carry on the work authorized under act of appropriation of June 25, 1910, which made available \$100,000—

```
<div type="message">  
<head type="main">MESSAGE.</head>  
<opener>  
<salute><i>To the Senate and House of Representatives:</i></salute>  
</opener>  
<p>I submit for the information of Congress the ...</p>
```

4.7 Exhibit 7

MISS TESMAN.

Yes, in order that Hedda may not be ashamed of me when we are walking in the street together.

TESMAN.

[Patting her under the chin.] You positively think of everything, Aunt Julie! *[Puts the hat on a chair close to the table.]* Now, look here, let us sit down here on this sofa and chat a little until Hedda comes.

[They sit down. She places her parasol on the settee.]

<sp>

<speaker>MISS TESMAN.</speaker>

<p>Yes, in order that Hedda may not be ashamed of me when we are walking in the street together.</p>

</sp>

<sp>

<speaker>TESMAN.</speaker>

<p><stage><i>Patting her under the chin.</i></stage> You positively think of everything, Aunt Julie!

<stage><i>Puts the hat on a chair close to the table.</i></stage>

Now, look here, let us sit down here on this sofa and chat a little until Hedda comes.</p>

<stage><i>They sit down. She places her parasol on the settee.</i></stage>

</sp>