

# **CDL TEI Base Encoding Guidelines**

**Nancy Kushigian**

## **CDL TEI Base Encoding Guidelines**

Nancy Kushigian

Copyright © 2006 University of California

**The encoding guidelines provided here are unedited working drafts produced by CDL's Structured Text Working Group. They should not be treated as final documents. Updated guidelines will be available in November 2004.**



# Table of Contents

Introduction .....	vi
1. General Instructions .....	1
1.1. Tag Sets .....	1
1.2. Encoding Practice .....	1
1.3. Element Names .....	2
1.4. Character Encoding .....	2
1.5. Extent of Encoding .....	3
1.6. Unique Identifier .....	3
2. Encoding Practice .....	4
2.1. Root Element .....	4
2.2. Document Header .....	4
2.3. Text Structure .....	5
2.3.1. Groups of Texts .....	6
2.4. Front Matter .....	7
2.4.1. Title Page .....	7
2.4.2. <b>&lt;docTitle&gt;</b> .....	8
2.4.3. <b>&lt;docAuthor&gt;</b> .....	8
2.4.4. <b>&lt;byline&gt;</b> .....	9
2.4.5. <b>&lt;docImprint&gt;</b> , <b>&lt;publisher&gt;</b> , <b>&lt;pubPlace&gt;</b> .....	9
2.4.6. <b>&lt;docDate&gt;</b> .....	9
2.4.7. <b>&lt;epigraph&gt;</b> .....	10
2.5. Divisions .....	10
2.6. Division Openers and Closers .....	11
2.6.1. <b>&lt;head&gt;</b> .....	11
2.7. Page Breaks and Milestones .....	13
2.7.1. <b>&lt;pb&gt;</b> .....	13
2.8. Arbitrary Containers and Segments .....	13
2.9. Font Changes .....	14
2.9.1. <b>&lt;hi&gt;</b> .....	14
2.9.2. <b>&lt;emph&gt;</b> .....	14
2.10. Foreign Words .....	14
2.11. Block Quotations .....	15
2.12. Verse .....	15
2.12.1. Divisions in verse. ....	15
2.12.2. Titles and Headings for Verse .....	15
2.12.3. Encoding Line Groups and Fragments .....	16
2.12.4. Lines of Verse and Indentation .....	16
2.12.5. Typographic Divisions within Poems .....	16
2.12.6. Closers and Trailers .....	16
2.13. Internal Links and Cross References .....	17
2.14. External Objects .....	17
2.15. Notes .....	18
2.15.1. Footnotes and Endnotes .....	19
2.16. Lists .....	20
2.16.1. Ordered Lists .....	20
2.17. Bibliographies .....	22
2.18. Figures & Formulas .....	22
2.19. Tables .....	23
2.20. Back Matte .....	23
3. Files and File-Naming .....	24
4. Quality Assurance .....	25
4.1. Validation .....	25
4.2. Best Practice Checking .....	25

4.3. Proofreading .....	25
Tag Library .....	26

# Introduction

The following base-encoding guidelines are not meant to be used alone, but rather should be used in the development of best practice documents for specific document types (i.e. manuscripts, books, oral histories, etc.). Elements or encoding designated "optional" in this document may be "required" by these text type guidelines. All STWG guidelines are meant to be used in conjunction with documentation for TEI P4. Where an issue is not directly addressed in these guidelines, full TEI documentation should be consulted.

# Chapter 1. General Instructions

## 1.1. Tag Sets

Use the full TEI XML DTD and include the prose, figures, and linking tag sets as shown below. The use of the CDL user extension files and the inclusion of the 'STWG.book' entity is the mechanism by which a text type DTD extensions is selected.

Entity declarations should use the full object filename, less extension. The SYSTEM identifier could either point at a publicly available URL or local system path. However, in the latter case, this path must be replicated in the file set submitted to CDL. Please use the 'GIF', 'JPEG', or 'PDF' notations. Texts **MUST** validate before being submitted to CDL.

```
<!DOCTYPE TEI.2 SYSTEM "tei2.dtd" [
<!ENTITY % TEI.XML "INCLUDE">
<!ENTITY % TEI.prose "INCLUDE">
<!ENTITY % TEI.figures "INCLUDE">
<!ENTITY % TEI.linking "INCLUDE">

<!ENTITY % TEI.extensions.ent SYSTEM 'CDL_base.ent'>
<!ENTITY % TEI.extensions.dtd SYSTEM 'CDL_base.dtd'>
<!ENTITY % CDL.book "INCLUDE">

<!ENTITY fig1 SYSTEM "figures/fig1.gif" NDATA GIF>
.
.
.]>
```

[P4: 3.3] [<http://www.tei-c.org/P4X/ST.html#STIN>]

If images are to be supplied in multiple resolutions, it will be necessary to encode this in the METS record for the text:

```
<fileGrp ID="figures">
  <fileGrp ID="fig1">
    <file ID="fig1-m" ADMID="image-rights" USE="med-res" MIMETYPE="image/gif">
      <FLocat LOCTYPE="URL"
        xlink:href="/dynaxml/data/cj/kt109nc2cj/figures/fig1.gif"/>
    </file>
    <file ID="fig1-h" ADMID="image-rights" USE="hi-res" MIMETYPE="image/gif">
      <FLocat LOCTYPE="URL"
        xlink:href="/dynaxml/data/cj/kt109nc2cj/figures/fig1_h.gif"/>
    </file>
  </fileGrp>
  ...
```

Please consult with the CDL Ingest team before constructing a multiple-resolution METS.

## 1.2. Encoding Practice

Markup practice should adhere strictly to the following guidelines, especially with regard to those elements, attributes and

attribute values that are marked as "REQUIRED". If a question arises that cannot be resolved with these instructions, the contributor should resort to the TEI Guidelines, version 4:

<http://www.tei-c.org/P4X/>

At all points in these guidelines the appropriate TEI Guideline section will be referenced using the following notation:

[P4: 11.2]

## 1.3. Element Names

Use 'camel case' element names only [i.e. `teiHeader`, NOT `teiheader` or `TEIHEADER`].

## 1.4. Character Encoding

Documents must be encoded using the Unicode Standard (UTF-8), and must include the encoding attribute in their XML declaration:

```
<?xml version="1.0" encoding="UTF-8"?>
```

Unicode characters can be incorporated into a document directly (à) or indirectly by representing them with numeric entity references. These numeric character entities can take either the decimal (&#224;) or hexadecimal forms (&#x00E0;). Characters can NOT be represented using named character entities (&grave;), with the exception of those specifically exempted in the XML 1.0 Specification, which MUST be used to avoid validation errors:

```
'<' Less Than      &lt;
> Greater Than    &gt;
& Ampersand      &amp;
```

If attribute values need to contain single or double quotes use the following character entities to avoid a parser error:

```
" Quotation Mark  &quot;
' Apostrophe      &apos;
```

### Example:

```
<name reg="Ol&apos; Yeller">
```

As part of the CDL ingest process, documents will be checked for the correct encoding and rejected if nonconforming characters or encodings are detected.

## 1.5. Extent of Encoding

CDL TEI Text Type Guidelines specify minimal encoding for particular types of texts. Normally, the <encodingDesc> element in the document header defines your encoding practices.

## 1.6. Unique Identifier

Every text will be assigned an ARK (Archival Resource Key) (i.e. 'ark:/13030/kt167nb66r') as a unique identifier by CDL, which should be used for file naming (see below).

# Chapter 2. Encoding Practice

## 2.1. Root Element

Each document should contain one and only one root element. The 'id' attribute is required and should contain the unique key of the ARK assigned to the text in question:

```
<TEI.2 id="kt5n39n99v">
```

## 2.2. Document Header

Header encoding should follow the practices described in detail in the *CDL Encoding Guidelines for the TEI Header* [<http://www.cdlib.org/inside/diglib/stwg/metadata>]. It is usually sufficient to follow the instructions there for the minimal header. However, if your project is going to depend on the TEI header as the principal source of metadata it is advisable to use the full header as a model. CDL searching depends on a crosswalk to Dublin Core for indexing. You will find instructions on which elements are essential for this process in the previously mentioned document. (*CDL Encoding Guidelines for the TEI Header* [<http://www.cdlib.org/inside/diglib/stwg/metadata>], Appendices A & B)

NOTE: I would like to stress the importance of including the <idno type="ARK"> element and <idno type="LOCAL"> (used to preserve the project's local object id). These are essential to maintaining the link between the document, it's persistent identifier (ARK), and its local ID.

Minimal Header Example:

```
<teiHeader type="CDL-TEI:BK">
  <fileDesc>
    <titleStmt>
      <title>The Opening of the Apartheid Mind :
        Options for the New South Africa :
        Electronic Version</title>
    <author>
      <name>Adam, Heribert</name>
    </author>
    <respStmt>
      <resp>and</resp>
      <name reg="Moodley, Koogila">Kogila Moodley</name>
    </respStmt>
    <respStmt>
      <resp>Text encoder:</resp>
      <name reg="Hastings, Kirk">Kirk Hastings</name>
    </respStmt>
  </titleStmt>
  <extent>ca. 816 Kb</extent>
  <publicationStmt>
    <publisher>California Digital Library</publisher>
    <pubPlace>Oakland, CA</pubPlace>
    <date value="2002-00-00">2002</date>
    <idno type="ARK">ark:/13030/ft958009mm</idno>
    <idno type="LOCAL">6178</idno>
  </publicationStmt>
  <seriesStmt>
```

```

    <title>UC Press Electronic Editions</title>
    <idno type="OTHER">http://www.ucpress.edu</idno>
  </seriesStmt>
  <sourceDesc>
    <biblFull>
      <titleStmt>
        <title>The Opening of the Apartheid Mind :
          Options for the New South Africa</title>
        <author>
          <name reg="Adam, Heribert">Heribert Adam</name>
        </author>
        <respStmt>
          <resp>and</resp>
          <name reg="Moodley, Koogila">Kogila Moodley</name>
        </respStmt>
      </titleStmt>
      <extent>xvi, 277 p. : map ; 23 cm.</extent>
      <publicationStmt>
        <publisher>University of California Press</publisher>
        <pubPlace>Berkeley, CA</pubPlace>
        <date value="1993-00-00">1993</date>
        <idno type="LCCN">92036443</idno>
      </publicationStmt>
    </biblFull>
  </sourceDesc>
</fileDesc>
</teiHeader>

```

[P4: 5.6] [<http://www.tei-c.org/P4X/HD.html#HD7>]

## 2.3. Text Structure

All information in the source document should be encoded within the overarching `<text>` element. The `<text>` element contains three sub-elements, the `<front>`, `<body>`, and `<back>`. Of these three, only `<body>` is required. These three `<text>` divisions are further subdivided into a hierarchy of divisions, beginning with a mandatory `<div1>` at the top of this hierarchy for `<body>` and `<back>`.

```

<TEI.2 id="ARK">
  <teiHeader> . . . </teiHeader>
  <text>
    <front> . . . </front>           OPTIONAL
    <body>                               REQUIRED
      <div1> . . . </div1>
    </body>
    <back>                               OPTIONAL
      <div1> . . . </div1>
    </back>
  </text>
</TEI.2>

```

[P4: 7.1] [<http://www.tei-c.org/P4X/DS.html#DSDIV>]

## 2.3.1. Groups of Texts

Groups of individual texts packaged together within a single document should be encoded using TEI's <group> element. Only use <group> when each individual text has its own title page. Groups of texts which do not have their own individual title pages should be encoded using TEI's standard <divn> element structure.

```
<TEI.2>
  <teiHeader type="text" status="new">
    <fileDesc>
      <titleStmt>
        <title/>
      </titleStmt>
      <publicationStmt>
        <p>x</p>
      </publicationStmt>
      <sourceDesc default="NO">
        <p>x</p>
      </sourceDesc>
    </fileDesc>
    <profileDesc>
      <langUsage default="NO">
        <language id="eng">English</language>
        <language id="spa">Spanish</language>
      </langUsage>
    </profileDesc>
  </teiHeader>
  <text>
    <front>
      <titlePage>
        <pb id="pi" n="[i]"/>
        <docTitle>
          <titlePart type="main">CONVENIO COLECTIVO ENTRE SAMUEL S. VENER COMPANY Y LA U
            CAMPESINOS DE AMERICA, AFL-CIO</titlePart>
          <titlePart type="main">COLLECTIVE AGREEMENT BETWEEN SAMUEL S. VENER COMPANY AN
            FARM WORKERS OF AMERICA-CIO</titlePart>
        </docTitle>
        <pb n="[ii]" id="pii"/>
      </titlePage>
    </front>
    <group>
      <text lang="eng">
        <front>
          <div1 type="index" org="uniform" part="N" sample="complete">
            <pb id="piii" n="[iii]"/>
            <head type="main">INDEX</head>
            <list type="simple">
              <item>ARTICLE 1: Recognition</item>
              <item> ... </item>
            </list>
          </div1>
        </front>
        <body>
          <div1 type="article" org="uniform" part="N" sample="complete">
            <head id="a1" type="main">ARTICLE 1: RECOGNITION</head>
            <p>A. The Company does hereby recognize the Union as the sole labor organiza
              representing all of the Company's agricultural employees (hereinafter call
              "workers") in the unit set forth in Agricultural Labor Relations Board's c
            <p> ... </p>
            <p>Executed and effective as of this 7th day of April, 1977.</p>
          </div1>
        </body>
      </text>
    </group>
  </text>
```

```

<text lang="spa">
  <front>
    <titlePage>
      <pb id="pia" n="[i]"/>
      <docTitle>
        <titlePart type="main">CONVENIO COLECTIVO S. VENER COMPANY Y LA UNION DE C
          AMERICA, AFL-CIO</titlePart>
        <pb id="piia" n="[ii]"/>
      </docTitle>
    </titlePage>
    <div1 org="uniform" part="N" sample="complete">
      <pb n="[iii]" id="piiaa"/>
      <head rend="underline">INDICE</head>
      <list type="simple">
        <item>ARTICULO 1 Reconocimiento</item>
        <item> ... </item>
      </list>
    </div1>
  </front>
  <body>
    <div1 type="article" org="uniform" part="N" sample="complete">
      <head type="main" id="als">
        <hi rend="underline">ARTICULO 1: RECONOCIMIENTO</hi>
      </head>
      <p>A. Por este medio la Compañía reconoce a la Unión como la única organizac
        que representa a todos los empleados agrícolas ...</p>
    </div1>
  </body>
  <back>
    <div1 type="appendix" org="uniform" part="N" sample="complete">
      <pb id="p28a" n="28"/>
      <head type="main" id="apas">
        <hi rend="underline">APENDICE "A"</hi>
      </head>
      <head type="subtitle">SUELDOS</head>
      <p> ... </p>
    </div1>
  </back>
</text>
</group>
</text>
</TEI.2>

```

NOTE: All of the texts within such a composite document share the same, single <teiHeader> at the very beginning of the file.

[P4: 7.3] [<http://www.tei-c.org/P4X/DS.html#DSGRP>]

## 2.4. Front Matter

Front matter will vary depending on the type of text. Most of what is described here applies to printed materials and may not be applicable to manuscripts.

### 2.4.1. Title Page

Do not encode a <titlePage> unless the document being encoded itself contains a formal title page. There is a decided lack of elements available in TEI for all of the various entities one typically finds on title pages. Title pages within source

documents may utilize a number of formatting peculiarities such as specific alignment, fonts, incidental images, etc. The encoder need not necessarily concern herself with reproducing this display exactly. It is often enough to convey to users that the document has a title page, what textual information it contains and the order in which it appears.

**Example:**

```
<titlePage>
  <docTitle>
    <titlePart type="main">The Opening of <lb/>the Apartheid Mind</titlePart>
    <titlePart type="subtitle">Options for the New South Africa</titlePart>
  </docTitle>
  <docAuthor><name>Heribert Adam</name> and
    <name>Kogila Moodley</name></docAuthor>
  <docImprint>
    <publisher>UNIVERSITY OF CALIFORNIA PRESS</publisher>
    <pubPlace>Berkeley · Los Angeles · London</pubPlace>
    <docDate>1993</docDate>
  </docImprint>
</titlePage>
```

## 2.4.2. <docTitle>

Encode the formal title, subtitle, and other subsidiary title parts within the required <docTitle>.

**Example:**

```
<docTitle>
  <titlePart type="main">Inventory of furniture and art.</titlePart>
</docTitle>
```

## 2.4.3. <docAuthor>

Record here authors and other names appearing on the title page responsible for the intellectual content of the document. Each <docAuthor> will generally render on its own line except when it appears within a <byline> element. Encoders may choose to encode multiple names within a single <docAuthor> if it is desired that they display on a single line, or may choose to repeat <docAuthor> if they should be displayed on separate lines. When encoding more than one name in a single <docAuthor>, encoders may optionally choose to further encode each name in a <name> sub-element.

Record the name as it appears on the title page. Do not attempt to reorder the name into catalog entry form or use the form of the name as it may appear in a name authority file. The reg attribute may be used to correlate a name to an authority.

**Example:**

```
<docAuthor>
  <name>Tom Jennings </name> and
  <name>Julia Hoffman, MD</name>
</docAuthor>
```

OR:

```
<docTitle>
  <titlePart type="main">Canine morphotypes and physiology</titlePart>
</docTitle>
<docAuthor>Tom Jennings </docAuthor>
<docAuthor>Julia Hoffman, MD</docAuthor>
```

#### 2.4.4. <byline>

Encode statements of responsibility in a <byline> element. <docAuthor>s are very frequently encoded within <byline> when they are associated with text ascribing a role or responsibility for the intellectual content of the work, e.g., "by."

Example:

```
<docTitle>
  <titlePart type="main">Canine morphotypes and physiology</titlePart>
</docTitle>
<byline>By <docAuthor>Tom Jennings </docAuthor> and
<docAuthor>Julia Hoffman, MD</docAuthor></byline>
```

#### 2.4.5. <docImprint>, <publisher>, <pubPlace>

Record publication information within <docImprint>. Use <pubPlace> and <publisher>, in any order and as often as necessary to record the place or places of publication and the name of the publisher or publishers respectively.

Example:

```
<docImprint>
  <pubPlace> Collinsport:</pubPlace>
  <publisher> Stoddard and Associates, 1993.</publisher>
</docImprint>
```

#### 2.4.6. <docDate>

Record copyright and publication dates within <docDate>. Include any associated text or symbols such as the word "copyright" or.

Example:

```
<docDate>&#xA9; 1941.</docDate>
```

## 2.4.7. <epigraph>

Quotations or bits that appear on the title page may simply be encoded as a segment or paragraph or they may be encoded within <epigraph>. Assign a rend attribute if desired to record specific renderings such as italics. Within <epigraph>, unattributed text may be recorded in a <q> element While attributed epigraphs may be surrounded with <cit>.

Example:

```
<epigraph>
  <q>The price we pay one day may<lb>make us weep.</q>
</epigraph>
```

OR:

```
<p>
  <q>The price we pay one day may make us weep.</q>
```

Example:

```
<epigraph rend="italic">
  <cit>
    <q>No man is an island, but some men are peninsulas.</q>
    <bibl>Joe Haskell</bibl>
  </cit>
</epigraph>
```

[P4: 7.5] [<http://www.tei-c.org/P4X/DS.html#DSTITL>]

## 2.5. Divisions

The document should use a hierarchical structure of numbered <divn> elements, with at least one <div1> in <body> and <back>. No <div> or <div0> elements are allowed. Do not surround an entire text or group with a <divn> if there are no smaller divisions within. All <divn> elements must have a unique 'id' attribute for purposes of navigation. If necessary these can be added automatically on ingest by CDL:

```
<div1 id="ch01">
  <div2 id="ss1.1">
    <div3 id="ss2.1">
      <div4 id="ss3.1">
```

**[P4: 7.1.2] [<http://www.tei-c.org/P4X/DS.html#DSDIV2>]**

The type attribute is required on all <divn> elements. Every attempt should be made to supply the most specific type possible for <divn> elements. CDL encoding guidelines for specific types of texts include sets of types appropriate to that genre. The following is a list that might be appropriate to for published books:

<divn type="sec">	section
volume	Volume
part	Part
chapter	Chapter
ss1...ss6	Sub-section level number
endnotes	End notes section
index	Back of the book index
dedication	Dedication
bibliography	Bibliography
appendix	Appendix
glossary	Glossary

## 2.6. Division Openers and Closers

### 2.6.1. <head>

The <head> element is used by the system for navigation purposes. Specific guidelines are supplied below regarding where headers may or may not appear. A head element should always be used where a display or where system-supplied navigation is desired. Heads may be supplied by the editor/encoder if they are not available in the text. Supplied heads should be enclosed in square brackets or some other convention expressly detailed in the editorial declaration. No type attribute is required by the system, although CDL encoding guidelines for specific types of texts will specify and require some type attributes.

Examples:

Title transcribed from text:

```
<head>Chapter 4. The Ghost Returns to Middlington Manor.</head>
```

Title supplied by editor/encoder:

```
<head>[Segment 2]</head>
```

Examples of what NOT to do:

In general, it is good practice to use the <head> tag for all major textual divisions. In order to support document navigation and display the following rules for the use of the <head> tag must be followed:

1. If any sibling of a <divn> at any level contains a head, then it must as well. Therefore, if any div1 uses a head, all div1's in the text must do so, etc.
2. If div's are left headless at a certain level, then any divs subordinate to them may not have heads.

Multiple <head> elements can be differentiated using the 'type' attribute (i.e. "subtitle" for subtitle)s:

```
<div1 id="ch01">
  <head type="main"> . . . </head>
  <head type="subtitle"> . . . </head>
```

The following example is wrong because one of the sibling <div1>s contains a <head> while none of the other <div1>s contain them:

```
<div1></div1>
<div1></div1>
<div1><head></head></div1>
<div1></div1>
```

The next example is wrong because all of the child <div2>s contain a <head> but the parent <div1> does not:

```
<div1>
  <div2><head></div2>
  <div2><head></div2>
</div1>
```

The last example is wrong because one of the <divn> descendants contains a <head> but none of its ancestors contain one.

```
<div1>
  <div2></div2>
  <div2></div2>
  <div2>
    <div3></div3>
    <div3>
      <div4>
        <head></head>
      </div4>
    </div3>
  <div2>
  <div2></div2>
</div1>
```

In fact, in this last example, ALL the <divn>s must contain a head! The single <div4> with a <head> has made it necessary for the entire tree to have <heads>.

## 2.7. Page Breaks and Milestones

### 2.7.1. <pb>

You must record a page break marking the beginning of each page of your source document. The n attribute may be used by the system to render page numbers.

```
<div1 type="chapter" n="I" id="ch01">
  <pb n="1" id="p1"/>
  <head>Introduction</head>
```

The <milestone> element may be used to mark boundaries between sections of text that are not page breaks or normal divisions. The unit attribute is required. One possible use would be to encode the decorative section breaks common to monographs:

```
<milestone unit="section" rend="decorative">
```

[P4: 6.9.3] [<http://www.tei-c.org/P4X/CO.html#COR55>]

## 2.8. Arbitrary Containers and Segments

Arbitrary Containers (<ab> and <seg>) may be used to contain text, groups of images or tables that are in positions where they wouldn't normally be valid. This can happen, for example, when it is necessary to encode a <note> within a closer or head. Use <seg> when you need to tag a phenomenon that occurs within major divisions, but use <ab> to tag phenomena that occur between major divisions (a list of valid locations for <seg> and <ab> can be found in the TEI Guidelines). Both elements must be assigned a type attribute. For instance, use <ab> rather than inserting extra <p>'s around non-paragraphs simply to allow the text to validate.

```
<ab type="illgrp">
  <figure id="fig001" entity="kt167nb66r_fig001.gif">
</ab>
```

Both elements must be assigned a type attribute.

NOTE: <ab> and <seg> should ONLY be used as a last resort, and not simply to allow a text to validate.

[P4: 14.3] [<http://www.tei-c.org/P4X/SA.html#SASE>]

## 2.9. Font Changes

### 2.9.1. <hi>

Record special typography with the <hi rend=""> tag. Use the required rend attribute to record the type of typography employed in the source document, e.g., rend="italic", rend="underline". Values for the rend attribute are described in encoding guidelines for various types of texts. Rend should NOT be used for editorial changes in typography unless the practice is documented in the encoding description.

### 2.9.2. <emph>

The <emph> element may be used instead of <hi> to distinguish typography that conveys emphasis rather than simply stylization. Thus, one could encode the following:

Example:

```
<hi rend= "largecaps">Once Upon a Time</hi> Chicken Little decided
to build a <emph rend="italics">very</emph> big house.
```

CDL supports the following 'rend' attribute values for <hi> & <emph>:

```
<hi rend="bold">
    italic
    mono
    roman
    smallcaps
    strikethrough
    subscript
    superscript
    underline
    Monospace
    Deemphasized
```

If the 'rend' attribute is omitted text will be rendered italic.

[P4: 6.3.2.2] [<http://www.tei-c.org/P4X/CO.html#COHQHE>]

## 2.10. Foreign Words

Use the foreign element to indicate a word or phrase that might require a different font or writing direction. The 'lang' attribute should contain the name of the appropriate UNICODE code chart. (see <lanfUsage>).

```
<foreign lang="Greek">&#904;&#905;&#906;&#907;&#908;</foreign>
```

[P4: 6.3.2] [<http://www.tei-c.org/P4X/CO.html#COHQH>]

## 2.11. Block Quotations

Simple block quotes may be recorded as:

```
<p rend="blockquote">It was seen from the beginning of the study . . . </p>
```

If your quote contains multiple paragraphs, use the <q> tag to enclose the entire quote and <p> do contain the individual paragraphs:

```
<q rend="blockquote">
  <p>It was seen from the beginning that the study . . . </p>
  . . .
</q>
```

[P4: 6.3.3] [<http://www.tei-c.org/P4X/CO.html#COHQQ>]

## 2.12. Verse

### 2.12.1. Divisions in verse.

Poems in the body of the text of a volume of poetry will be tagged using <div1> for the main structural element, with <div2> tags (and so on) for subordinate sections within the poem.

The most common attributes for poems will be:

```
<div1 type="poem">
  "sonnet"
  "drama"
```

See encoding guidelines for books, rare books, and manuscripts for more complete discussions of poetry and for recommended attribute values.

### 2.12.2. Titles and Headings for Verse

Use the <head> tag for all titles, subtitles, etc. For anything but the main title, use the type attribute to note its function:

```
<head type="subtitle">
<head type="dedication">
```

## 2.12.3. Encoding Line Groups and Fragments

Poetry is generally encoded by using the line group `<lg>` element and the line `<l>` element. The `<lg>` tag is used around groups of lines that carry poetic structure, such as verses, quatrains, etc., which can be specified in a type attribute as in the example below:

```
<div1 type="poem">
  <lg type="stanza">
    <l>Twinkle twinkle little star,</l>
    <l>How I wonder what you are,</l>
    <l>Up above the world so high,</l>
    <l>Like a diamond in the sky,</l>
    <l>Twinkle twinkle little star,</l>
    <l>How I wonder what you are.</l>
  </lg>
</div1>
```

The attribute 'type' in `<lg>` is optional. However encoding guidelines for types of texts will specify recommended attributes (common ones are "stanza," "verse paragraph," "couplet," "quatrain," etc.)

For line groups that carry no poetic structure (e.g. fragments) use:

```
<lg type="fragment">
```

## 2.12.4. Lines of Verse and Indentation

Lines of verse use the `<l>` tag. Lines that are indented are noted using the `rend` attribute:

```
<l rend="indentn"> where n=the number of indentation units desired.
```

The attributes 'rend' and 'n' are optional.

## 2.12.5. Typographic Divisions within Poems

Typographic division within poems, such as a line of asterisks or periods or a line may be encoded as `<milestone>` with attribute values of `unit="typography"` and `n="*****"` or `n="....."` (where the value of n is the typographic character used to divide the line groups.)

## 2.12.6. Closers and Trailers

Occasionally, a poem will end with a date, or some other type of information not part of the poem. This information may be encoded using the `<closer>` tag, as for example:

```
<closer>1876.</closer>
```

For more detailed discussions of verse encoding, see the text type guidelines and the TEI Guidelines.

## 2.13. Internal Links and Cross References

References will be encoded using the <ref> element and are required to have both a target and type attribute to indicate the nature of the target. In the case of a published book the following values might apply:

```
<ref target="#" type="fnoteref"> footnotes
    noteref endnotes
    pageref page breaks
    citeref bibliographic citations
    figref figures
    tableref tables
    formularef formulas
    secref any div
```

Example:

```
<ref id="bkd0e131" target="d0e131" type="noteref">1</ref>
<note id="d0e131" corresp="bkd0e131" place="end" n="1">
.
.
.
</note>
```

[P4: 6.6] [<http://www.tei-c.org/P4X/CO.html#COXR>]

## 2.14. External Objects

We have upgraded the <xref> and <xptr> elements by adding the 'href' attribute for external URIs, as suggested in the TEI guidelines. Now the 'doc' attribute can contain a local object's entity name and 'href' can be used for URLs or ARKs, which are not legal entity names. The 'type' attribute can now be used to indicate the link or object type, the 'rend' attribute to define how the object should be linked to, and the 'from' and 'to' attributes to point at a specific parts of the digital object. 'from' and 'to' usually target 'id' attributes on structural elements. The following set of attribute values are supported:

```
<xref doc="ENTITY" href="ARK" type="obj" rend="new" from="" to="">TEXT</xref>
xptr          URL          mets          replace
                url          embed
                pdf          none
                sound
                video
                stream
```

External Link Examples:

URL:

```
<xref href="http://www.cdlib.org" type="url" rend="new">
```

Result: new window displaying referenced URL.

CDL digital object:

```
<xref href="ark:/13030/kt5n39n99v" type="obj" rend="replace" from="ch02">
```

Result: document view replaced with chapter 2 of referenced ARK.

PDF document:

```
<xref doc="kt167nb66r_ch19.pdf" type="pdf" rend="new">
```

Result: new window displaying a PDF of chapter 19.

If a more complicated selection of derivatives needs to be offered the user, I would suggest making use of the `<structMap>` in the METS record. Where 'from' and 'to' would point to 'id' attributes in the referenced METS:

```
<xref href="ark:/13030/kt5n39n99v" type="mets" rend="embed" from="seg6">
```

Result: all objects under the 'seg6' portion of the `<structMap>` are embedded in the document.

[P4: 14.2] [<http://www.tei-c.org/P4X/SA.html#SAXR>]

## 2.15. Notes

For notes use the `<note>` element and the 'place' attribute to indicate whether it's an "end", "foot" or "inline" note.

Supported note "place" attributes are:

end

note appears at end of chapter, part, or volume.

foot

note appears at foot of page.

inline

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

Notes without the place attribute will be considered in-line. A footnote can be placed in the text immediately following the reference to it. Endnotes should be placed exactly as they appear in the printed book, in a separate <divn> if necessary. Note numbers are encoded as the value of the 'n' attribute and should not be included in the text. Reference dingbats, crosses, daggers, and the like may be recorded as Unicode symbols in either n="\*" or in the content for <ref>\*</ref>. If a note is referenced by a <ref>, it must contain a unique 'id'. Place any note text in <p>s.

```
<note id="n1_ch13" place="foot" n="*">
  <p>In the literature, some confusion exists . . . </p>
</note>
```

## 2.15.1. Footnotes and Endnotes

Encode footnotes (those references, notes and citations appearing at the bottom of the page) where they are referenced and endnotes (those appearing at the end of a chapter, section, or other significant textual division) at the point in the text where they actually appear. Standardizing the encoding of such notes affords greater latitude in how stylesheets may render them. In a web interface, notes may appear in popup windows, while in printed versions they could be rendered as footnotes or endnotes. Use the n attribute of <note> to record the footnote number. Also assign a unique id to each note. A 'place' attribute with the value "foot" or "end" is required for every footnote or endnote. If you choose to encode a footnote that is clearly bibliographic in nature, you may enclose it within the TEI <bibl> element, encoding the author or authors as <author>, italicized or underlined titles as <title>, dates as <date>, and references to a page number, span of page numbers, or chapters as <biblScope>:

Examples with accompanying references:

INLINE:

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

ENDNOTE:

```
<ref id="ref1" target="en1" type="noteref"/>
<note id="en1" corresp="ref1" n="1" place="end">
  <p>We explain below why we use the uncommon term collection instead
  of the expected set. Our usage corresponds to the aggregate of
  many mathematical writings and to the sense of class found in
  older logical writings.</p>
```

</note>

FOOTNOTE:

<p>Collections are ensembles of distinct entities or objects of any sort.<note id="fn1" n="\*" place="foot"><p>We explain below why we use the uncommon term collection instead of the expected set. Our usage corresponds to the aggregate of many mathematical writings and to the sense of class found in older logical writings.</p></note> The elements...</p>

[P4: 6.8.1] [<http://www.tei-c.org/P4X/CO.html#CONONO>]

## 2.16. Lists

Use the TEI list scheme to encode lists. Use the type attribute of <list> to record the type of list appearing in the document. Valid type attributes are:

### **ordered**

Numbered lists

### **bulleted**

Marked or bulleted lists

### **simple**

Unmarked or unnumbered lists

### **gloss**

Definition lists consisting of a <label> and an <item>

Nest lists as appropriate. Use a list of type "gloss" for so-called "definition" lists consisting of a term (often highlighted) followed by a definition or expansion of the term. Gloss-type lists should be used for chronological lists (lists consisting of a date followed by an event, such as within biographies).

### 2.16.1. Ordered Lists

Encode lists which include sequential markers, numbers or letters as <list type="ordered">. Typically such sequential markers will begin with the numeral 1 and increment uniformly for each item in the list. However other numbering schemes may also be encountered. For uniformly increasing lists which use alternative ordering methods, supply an appropriate rend attribute from the following list:

#### **arabic**

1., 2., 3., etc.

#### **upperalpha**

A., B., C., etc.

#### **loweralpha**

a., b., c., etc.

**upperroman**

I., II., III., etc.

**lowerroman**

i., ii., iii., etc.

**supplied**

For non-standard enumerations encoded within each <item>'s n attribute.

Use of the "arabic" rend value is optional since it is the default style.

**Example:****Departments**

- A. English
- B. History
- C. Biology
- D. Political Science

```
<list type="ordered" rend="upperalpha">  
  <head>Departments</head>  
  <item>English</item>  
  <item>History</item>  
  <item>Biology</item>  
  <item>Political Science</item>  
</list>
```

For lists which use a style of enumeration not available in this list, or use a non-sequential or otherwise non-standard method of enumeration, the specific mark of numeration may be supplied in an explicit n attribute within each individual <item> element. In such cases, set the <list>'s rend attribute to "supplied."

**Example:**

- 1. Food and supplies
- 2. Medicine
- 3. Fuel
- 5. Fuel storage containers
- 6. Radios

```
<list type="ordered" rend="supplied">  
  <item n="1">Food and supplies</item>
```

```

    <item n="2">Medicine</item>
    <item n="3">Fuel</item>
    <item n="5">Fuel storage containers</item>
    <item n="6">Radios</item>
</list>

```

Note that if any one <item> contains an explicit n attribute, all <item> elements immediately within the <list> must also contain an explicit n attribute, even if the remaining items conform to the standard enumerative conventions.

Never encode the sequential marker within the text of the <item> element. Such encoding will usually result in two duplicate numbers appearing before each <item> in the list. Either allow the rendering software to supply the number or move the number into the <item>'s n attribute.

[P4: 6.7] [<http://www.tei-c.org/P4X/CO.html#COLI>]

## 2.17. Bibliographies

In bibliographic citations, the <title> element must make use of the 'level' attribute from which the formatting will be derived. Other <bibl> sub-elements can be used, but are not required. Use <listBibl> to contain a set of <bibl>s. The <note> tag is illegal in <bibl>, but uncontained text is allowed, so you can include any bibliographical annotation directly within <bibl> without containing it within another element.

```

<listBibl>
  <bibl id="bib010_ch02">
    <author>Johnson, Douglas W.</author>
    <date>1919</date>.
    <title level="m">Shore processes</title>.
    <pubPlace>New York</pubPlace>,
    <publisher>Wiley & Sons</publisher>,
    <biblScope type="pages">584 pp.</biblScope>,
    <date>1919</date>.
  </bibl>
</listBibl>

```

[P4: 6.10.1] [<http://www.tei-c.org/P4X/CO.html#COBITY>]

## 2.18. Figures & Formulas

Figures, graphs, and photographs should be encoded using the <figure> element. The 'rend' attribute is required and is used to indicate whether the image is in-line (part of a string of text), a separate 'block', or a 'popup' (linked to a higher resolution version). The complete image name assigned to an image should also be used as the entity attribute for the <figure> element that references it. Figure captions may be encoded in the <head> element with the attribute type="caption", or are sometimes encoded using <seg type="caption"> within <figDesc>

```

<figure id="fig001" entity="kt167nb66r_fig001.gif" rend="block">
  <head type="caption">Bottom topography in the South Atlantic Ocean.</head>
</figure>

```

[P4: 22.3] [<http://www.tei-c.org/P4X/FT.html#FTGROV>]

## 2.19. Tables

Use the XHTML 1.1 table module rather than the TEI default table scheme for all tables and encode as much as possible for correct display in both Netscape and I.E. on the Windows and Mac platforms. Take care to encode definition lists as `<list type="gloss">` when encountered (these can be confused for two-column tables).

```
<table id="tab001">
  <caption>PERCENTAGES OF THE EARTH'S SURFACE</caption>
  <colgroup span="3">
    <col align="right" span="1"/>
    <col align="char" char="." span="1"/>
    <col align="char" char="." span="1"/>
  </colgroup>
  <thead>
    <tr>
      <th>Latitude</th>
      <th>%</th>
      <th>Cumulative %</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <td>40 N 30 W</td>
      <td>8.68</td>
      <td>8.68</td>
    </tr>
  </tbody>
</table>
```

[P4: 22.1] [<http://www.tei-c.org/P4X/FT.html#FTTAB>]

## 2.20. Back Matte

The optional `<back>` element may contain any number of `<divn>` elements. Back matter may include advertisements, indexes, bibliographies, appendices and other matter. Guidelines for encoding back matter is specified in the guidelines for various types of text.

## Chapter 3. Files and File-Naming

Each document should be encoded as a single XML file using the following naming convention:

ARK.xml

ARK is the unique key of the assigned ark (i.e. kt167nb66r.xml)

For ease of ingest by CDL you may choose to use the following conventions for images, PDFs and other associated content:

ARK\_NAME.EXTENSION

ARK is the unique key of the assigned ark, NAME is the result of whatever internal naming convention you would like to apply to individual files, and EXTENSION is the normal file format extension ('.gif', '.jpg', or '.pdf'.)

kt167nb66r\_fig002.gif

To facilitate the preview and ingest of TEI objects, the location specified by the ENTITY declaration for images and other associated files must point at an accessible URL.

```
<!ENTITY fig001 SYSTEM "http://www.server.domain/figures/fig001.gif" NDATA GIF>
...
<figure id="fig001" entity="fig001" rend="block">
```

If at all possible, two derivative images should be created for each plate, figure, graphic, or other pictorial element that appears as a discrete element. One of these should be at web resolution (72 ppi) and the same size as the figure in the printed text. The other image should be higher resolution (300 ppi) and again original size, but not exceeding 768 pixels in width. In-line images, such as formulas need only the low-resolution version. When necessary, images should be cropped and flipped for proper orientation for Web display. We are agnostic about the GIF and JPEG formats. The master version of the image (usually a TIFF) does not need to be submitted to CDL. However, if your project is interested in preserving it for future use, the CDL Preservation Repository is an appropriate place.

# Chapter 4. Quality Assurance

## 4.1. Validation

All texts will be validated on ingest and rejected if errors are detected.

## 4.2. Best Practice Checking

In addition to validation against the supplied DTDs, documents will be checked for conformance to the appropriate Best Practice Guidelines using the Schematron assertion language. Users can check their documents against this standard using the CDL Text Preview page

<http://texts.cdlib.org/dynaxml/preview.html>

## 4.3. Proofreading

Proofing the actual text of submitted documents is the responsibility of the contributor. At the very least texts should be spot-checked. If the project warrants it, documents should be proofread by a professional using the CDL Text Preview page:

<http://texts.cdlib.org/dynaxml/preview.html>

# Tag Library

Below you will find a brief description of every element supported under the CDL Standard for Printed Books and their attributes. Attribute value definitions take one of the following forms:

<b>ENTITY</b>	Entity Name defined in an entity declaration (<!ENTITY fig1 SYSTEM "fig1.gif" NDATA GIF>)
<b>ID</b>	Unique ID
<b>ID REFERENCE</b>	Reference to an existing ID
<b>TEXT</b>	Unrestricted text
<b>URI</b>	Uniform Resource Indicator
<b>(OPTION1   OPTION2)</b>	A set list of optional values from which the encoder must choose.

In addition each element and attribute is declared REQUIRED, RECOMMENDED, or OPTIONAL.

## <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.

### Attributes:

type	(illgrp   tblgrp   text)	REQUIRED
------	--------------------------	----------

See Also *Arbitrary Containers* .

## <address>

Contains a postal or other address, for example of a publisher, an organization, or an individual.

## <addrLine>

Contains one line of a postal or other address.

## <author>

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

See Also *Bibliographies* .

## <availability>

Supplies information about the availability of a text, for example any restrictions on its use or distribution, its copyright status, etc.

See Also *Document Header* .

## <back>

Back matter. Contains any appendixes, etc. following the main part of a text.

See Also *Text Structure* .

## <bibl>

Bibliographic citation. Contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

### Attributes:

id	ID	OPTIONAL
----	----	----------

---

corresp	ID REFERENCE	OPTIONAL
---------	--------------	----------

See Also *Bibliographies* .

**<biblFull>**

Contains a fully-structured bibliographic citation, in which all components of the TEI file description are present.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<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.

**Attributes:**

type	TEXT	OPTIONAL
------	------	----------

See Also *Bibliographies* .

**<body>**

Text body. Contains the whole body of a single unitary text, excluding any front or back matter.

See Also *Text Structure* .

**<byline>**

Contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.

See Also *Division Openers and Closers* .

**<catDesc>**

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

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<category>**

Category. Contains an individual descriptive category, possibly nested within a superordinate category, within a user-defined taxonomy.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<change>**

Summarizes a particular change or correction made to a particular version of an electronic text which is shared between several researchers.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<cit>**

A quotation from some other document, together with a bibliographic reference to its source.

See Also *Quotations* .

**<classDecl>**

Classification declarations. Contains one or more taxonomies defining any classificatory codes used elsewhere in the text

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<date>**

Contains a date in any format. The content of 'value' must follow the ISO 8601:2000 5.2.1 date format (yyyy-mm-dd)

**Attributes:**

value	(yyyy-mm-dd)	OPTIONAL
-------	--------------	----------

**<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.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Division Openers and Closers* .

**<div1-7>**

Level 1-7 text divisions. Used to encode the structural subdivisions of the front, body, or back of a text.

**Attributes:**

id	ID	REQUIRED
n	TEXT	OPTIONAL
type	TEXT	REQUIRED

See Also *Divisions* .

**<docAuthor>**

Document author. Contains the name of the author of the document, as given on the title page.

See Also *Title Page* .

**<docDate>**

Document date. Contains the date of a document, as given (usually) on a title page.

See Also *Title Page* .

**<docEdition>**

Document edition. Contains an edition statement as presented on a title page of a document.

**Attributes:**

id ID OPTIONAL

See Also *Title Page* .

**<docImprint>**

Document imprint. Contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.

See Also *Title Page* .

**<docTitle>**

Document title. Contains the title of a document, including all its constituents, as given on a title page.

See Also *Title Page* .

**<edition>**

Edition. Describes the particularities of one edition of a text.

See Also *Bibliographies* .

**<editionStmt>**

Edition statement. Groups information relating to one edition of a text.

**Attributes:**

id ID OPTIONAL

See Also *Document Header* .

**<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.

See Also *Bibliographies* .

**<editorialDecl>**

Editorial practice declaration. Provides details of editorial principles and practices applied during the encoding of a text.

**Attributes:**

id ID OPTIONAL

See Also *Document Header* .

**<emph>**

Emphasized. Marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.

**Attributes:**

rend (bold | italic | mono | roman | smallcaps | strikethrough | subscript | superscript | underline) OPTIONAL

See Also *Font Changes* .

**<encodingDesc>**

Encoding description. Documents the relationship between an electronic text and the source or sources from which it was derived.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<epigraph>**

Epigraph. Contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.

See Also *Division Openers and Closers* .

**<extent>**

Describes the approximate size of the electronic text as stored on some carrier medium, specified in any convenient units.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<figure>**

Indicates the location of a graphic, illustration, or figure.

**Attributes:**

id	ID	OPTIONAL
entity	ENTITY	REQUIRED
rend	(block   hide   inline   popup(ENTITY))	REQUIRED

See Also *Figures* .

**<fileDesc>**

File Description. Contains a full bibliographic description of an electronic file.

See Also *Document Header* .

**<foreign>**

Identifies a word or phrase as belonging to some language other than that of the surrounding text. The value of 'lang' should be a UNICODE code chart name (e.g. Greek, Hebrew, etc.)

**Attributes:**

lang	ID REFERENCE	REQUIRED
------	--------------	----------

See Also *Foreign Words* .

**<formula>**

Contains a mathematical or other formula.

**Attributes:**

id	ID	REQUIRED
notation	(mathML   TeX)	OPTIONAL
rend	(block   inline)	REQUIRED

See Also *Figures & Formulas* .

**<front>**

Front matter. Contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.

See Also *Front Matter* .

**<funder>**

Funding body. Specifies the name of an individual, institution, or organization responsible for the funding of a project or text.

**Attributes:**

id ID OPTIONAL

See Also *Document Header* .

**<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.

See Also *Groups of Texts* .

**<head>**

Heading. Contains any heading, for example, the title of a section, or the heading of a list or glossary.

**Attributes:**

type TEXT OPTIONAL

See Also *Division Openers and Closers* .

**<hi>**

Highlighted. Marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.

**Attributes:**

rend (bold | italic | mono | roman | smallcaps | strikethrough | subscript | superscript | underline) OPTIONAL

See Also *Font Changes* .

**<idno>**

Identifying number. Supplies any standard or non-standard number used to identify a bibliographic item.

**Attributes:**

type ( ARK | ISBN | ISSN | LCCN | LOCAL REQUIRED | OTHER )

See Also *Document Header* .

**<imprint>**

Groups information relating to the publication or distribution of a bibliographic item.

**Attributes:**

id ID OPTIONAL

See Also *Bibliographies* .

**<item>**

Contains one component of a list.  
See Also *Lists* .

**<keywords>**

Keywords. Contains a list of keywords or phrases identifying the topic or nature of a text.

**Attributes:**

id	ID	OPTIONAL
scheme	ID REFERENCE	OPTIONAL

See Also *Document Header* .

**<l>**

Verse line. Contains a single, possibly incomplete, line of verse.

**Attributes:**

n	TEXT	OPTIONAL
rend	(indent1   indent2   indent3   indent4   indent5   indent6   indent7   indent8   indent9   indent10)	OPTIONAL

See Also *Lines of Verse* .

**<label>**

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

**<language>**

Characterizes a single language or sub-language used within a text. The 'id' attribute should use a UNICODE code chart name (e.g. Greek, Hebrew, etc.)

**Attributes:**

id	ID	REQUIRED
----	----	----------

See Also *Document Header* .

**<langUsage>**

Language usage. Describes the languages, sub-languages, registers, dialects etc. represented within a text.  
See Also *Document Header* .

**<lb/>**

Line break. Marks the start of a new (typographic) line in some edition or version of a text.  
See Also *Milestones* .

**<lg>**

Line group. Contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.

**Attributes:**

type	(couplet   paragraph   quatrain   stanza   verse)	OPTIONAL
------	------------------------------------------------------	----------

See Also *Line Groups and Fragments* .

**<list>**

Contains any sequence of items organized as a list.

**Attributes:**

rend	(arabic   upperalpha   loweralpha   upperroman   lowerroman   supplied)	OPTIONAL
type	(bulleted   gloss   ordered   simple)	REQUIRED

See Also *Lists* .

**<listBibl>**

Citation list. Contains a list of bibliographic citations of any kind.

See Also *Bibliographies* .

**<milestone >**

Marks the boundary between sections of a text, as indicated by changes in a standard reference system.

**Attributes:**

id	ID	REQUIRED
rend	TEXT	OPTIONAL
unit	TEXT	REQUIRED

See Also *Milestones* .

**<monogr>**

Monographic level. Contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object).

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Bibliographies* .

**<name>**

Name, proper noun. Contains a proper noun or noun phrase.

**Attributes:**

reg	TEXT	OPTIONAL
type	(personal   place)	OPTIONAL

**<note>**

Contains a note or annotation.

**Attributes:**

corresp	ID	OPTIONAL
id	ID REFERENCE	REQUIRED
n	TEXT	OPTIONAL
place	(end   foot   inline)	OPTIONAL

See Also *Notes* .

**<p>**

Paragraph. Marks paragraphs in prose.

**Attributes:**

rend	(blockquote   center   hang   indent   noindent   left   right)	OPTIONAL
------	--------------------------------------------------------------------	----------

**<pb>**

Page break. Marks the boundary between one page of a text and the next in a standard reference system.

**Attributes:**

id	ID	REQUIRED
n	TEXT	OPTIONAL

See Also *Milestones* .

**<principal>**

Principal researcher. Supplies the name of the principal researcher responsible for the creation of an electronic text.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<profileDesc>**

Text-profile description. Provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sub-languages used, the situation in which it was produced, the participants and their setting.

See Also *Document Header* .

**<projectDesc>**

Project description. Describes in detail the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<ptr>**

Defines a pointer to another location in the current document in terms of one or more identifiable elements.

**Attributes:**

id	ID	OPTIONAL
target	ID REFERENCE	REQUIRED
type	TEXT	REQUIRED

See Also *Internal Links and Cross References* .

**<publicationStmt>**

Publication statement. Groups information concerning the publication or distribution of an electronic or other text.  
See Also *Document Header* .

**<publisher>**

Provides the name of the organization responsible for the publication or distribution of a bibliographic item.  
See Also *Title Page* .

**<pubPlace>**

Contains the name of the place where a bibliographic item was published.  
See Also *Title Page* .

**<q>**

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

**Attributes:**

rend	(blockquote)	OPTIONAL
------	--------------	----------

See Also *Quotations* .

**<ref>**

Defines a reference to another location in the current document, in terms of one or more identifiable elements, possibly modified by additional text or comment.

**Attributes:**

id	ID	OPTIONAL
target	ID REFERENCE	REQUIRED
type	TEXT	REQUIRED

See Also *Internal Links and Cross References* .

**<refsDecl>**

References declaration. Specifies how canonical references are constructed for this text.

**Attributes:**

id	ID	OPTIONAL
doctype	TEXT	OPTIONAL

See Also *Document Header* .

**<resp>**

Contains a phrase describing the nature of a person's intellectual responsibility.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<respStmt>**

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

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<revisionDesc>**

Revision description. Summarizes the revision history for a file. Attributes

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<seg>**

Arbitrary segment. Contains any arbitrary phrase-level unit of text.

**Attributes:**

type	(illgrp   tblgrp   text)	REQUIRED
------	--------------------------	----------

See Also *Arbitrary Containers* .

**<series>**

Series information. Contains information about the series in which a book or other bibliographic item has appeared.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<seriesStmt>**

Series statement. Groups information about the series, if any, to which a publication belongs.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<sourceDesc>**

Supplies a bibliographic description of the copy text(s) from which an electronic text was derived or generated.

See Also *Document Header* .

**<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.

**<sponsor>**

Specifies the name of a sponsoring organization or institution.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<table>**

Contains text displayed in tabular form, in rows and columns. Can contain the following elements: caption, td, th, tr, col, colgroup, tbody, thead, tfoot.

See Also *Tables* .

**<taxonomy>**

Taxonomy. Defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<TEI.2>**

TEI document. Contains a single TEI-conformant document, comprising a TEI header and a text. The value of 'id' should be the unique key of the ARK assigned to the text.

**Attributes:**

id	ID	REQUIRED
----	----	----------

See Also *Root Element* .

**<teiHeader>**

TEI Header. Supplies the descriptive and declarative information making up an 'electronic title page' prefixed to every TEI-conformant text.

See Also *Document Header* .

**<text>**

Contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, or a dictionary.

See Also *Text Structure* .

**<textClass>**

Text classification. Groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.

**Attributes:**

id	ID	OPTIONAL
----	----	----------

See Also *Document Header* .

**<title>**

Contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles.

**Attributes:**

level	(a   m   j   s   u)	OPTIONAL
type	(main   subtitle   alternate   abbreviated)	OPTIONAL

See Also *Bibliographies* .

**<titlePage>**

Title page. Contains the title page of a text, appearing within the front or back matter.  
See Also *Title Page* .

**<titlePart>**

Title part. Contains a subsection or division of the title of a work, as indicated on a title page.

**Attributes:**

type (main | subtitle | alternate | abbreviated) OPTIONAL

See Also *Title Page* .

**<titleStmt>**

Title statement. Groups information about the title of a work and those responsible for its intellectual content.  
See Also *Document Header* .

**<xptr>**

Extended pointer. Defines a pointer to another location in the current document or an external document. NOTE: The value of 'href' can be either a URL (e.g. <http://texts.cdlib.org/xtf/servlet/dynaXML?docId=ft958009mm>) or CDL ARK (e.g. <ark:/13030/ft958009mm>)

**Attributes:**

doc	ENTITY	OPTIONAL
href	URI	OPTIONAL
type	(mets   obj   pdf   sound   stream   url   video)	REQUIRED
rend	(embed   new   none   replace)	REQUIRED
from	ID	OPTIONAL
to	ID	OPTIONAL

See Also *External Objects* .

**<xref>**

Extended reference. Defines a reference to another location in the current document, or an external document, using an extended pointer notation, possibly modified by additional text or comment.

**Attributes:**

doc	ENTITY	OPTIONAL
href	ARK or URL	OPTIONAL
type	(mets   obj   pdf   sound   stream   url   video)	REQUIRED
rend	(embed   new   none   replace)	REQUIRED
from	ID	OPTIONAL
to	ID	OPTIONAL

See Also *External Objects* .