

California Digital Library TEI Lite Best Practice Guidelines for Encoding Basic Texts

California Digital Library Structured Text Working Group

California Digital Library TEI Lite Best Practice Guidelines for Encoding Basic Texts

California Digital Library Structured Text Working Group
Copyright © 2006 University of California

The encoding guidelines provided here are unedited working drafts produced by the 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
Using These Guidelines	vii
1. General Instructions	1
1.1. File Management and ARKs	1
1.1.1. Naming	1
1.1.2. Associated Content Files	1
1.1.3. Image Files	1
1.2. Invoking the CDL TEI Lite DTD	2
1.3. Case Sensitivity	3
1.4. Character Encoding	3
1.5. Hyphenation	4
1.6. Metadata Encoding and Transmission Standard (METS) Record	4
2. Encoding Practice	5
2.1. Root Element	5
2.1.1. <TEI.2>	5
2.2. Document Header	5
2.2.1. <teiHeader>	5
2.2.2. <idno>	6
2.3. Text Structure	6
2.3.1. <text>	6
2.4. Front Matter	7
2.4.1. <titlePage>	7
2.5. Document Body	9
2.5.1. <body>	9
2.6. Back Matter	9
2.6.1. <back>	9
2.7. Divisions	10
2.7.1. <divn>	10
2.8. Division Headings	10
2.8.1. <head>	10
2.9. Paragraphs	11
2.9.1. <p>	11
2.10. Page Breaks and Milestones	12
2.10.1. <pb>	12
2.10.2. <lb>	12
2.10.3. <milestone>	13
2.11. Typographical Phenomena and Formatting	13
2.11.1. <hi>	13
2.11.2. Nested <hi> Tags	14
2.11.3. <emph>	14
2.11.4. Alignment and Indention	15
2.12. Language Shifts	15
2.12.1. <foreign>	15
2.13. Quotations	16
2.13.1. <quote>	16
2.13.2. <cit>	17
2.14. Verse	17
2.14.1. <divn> in Verse	17
2.14.2. <head> in Verse	17
2.14.3. <l>	18
2.14.4. <lg>	18
2.15. Notes	19
2.15.1. <note>	19

2.15.2. In-line notes	20
2.15.3. Footnotes	20
2.15.4. Endnotes	20
2.16. Lists	21
2.16.1. <list>	21
2.16.2. Standard Ordered Lists	21
2.16.3. Non-standard Ordered Lists	22
2.16.4. <label>	23
2.17. Bibliographies	23
2.17.1. <bibl>, <listBibl>	23
2.17.2. <title> levels	23
2.17.3. <note> in Bibliographic Citations	24
2.18. Internal Links and Cross References	24
2.18.1. <ref>	24
2.19. External Objects	25
2.19.1. <xref>	25
2.20. Graphic Elements	27
2.20.1. <table>	27
2.20.2. <figure>	27
2.20.3. Formulas	28
2.21. Arbitrary Containers and Segments	29
2.21.1. <seg>	29
2.21.2. <ab>	29
3. Quality Assurance	30
3.1. Validation	30
3.2. Best Practice Checking	30
3.3. Proofreading	30
Tag Library	31

Introduction

This document is part of a collection of best practice guidelines established by the California Digital Library's Structured Text Working Group for encoding electronic texts. The guidelines provide best practices for marking up XML documents in accordance with the Textual Encoding Initiative's *TEI P4: Guidelines for Electronic Text Encoding and Interchange* (TEI P4). All projects submitting text documents to the CDL must follow the CDL TEI best practices in order to produce files that may be automatically ingested and distributed by the CDL. There are four separate but related guidelines available, each geared toward a specific type of text, each accompanied by a specific DTD:

- *California Digital Library TEI Lite Best Practice Guidelines for Encoding Basic Texts* [/stwg/lite/]
- *California Digital Library TEI Best Practice Guidelines for Encoding Printed Books* [/stwg/book/]
- *California Digital Library TEI Best Practice Guidelines for Encoding Manuscripts* [/stwg/ms/]
- *California Digital Library TEI Best Practice Guidelines for Encoding Oral Histories* [/stwg/oh/]

All of the above guidelines also require projects to consult the CDL's separate, universal set of guidelines for creating a TEI header: *California Digital Library Best Practice Guidelines for Encoding TEI Headers* [/stwg/metadata/]

These documents assume that readers are already familiar with the basics of XML and TEI P4 and are only seeking guidance as to how to apply them to specific cases. In other words, the best practices guidelines are not exhaustive instructions on XML nor the TEI. Not every element nor attribute available through a particular CDL TEI DTD is discussed in-depth, although a complete list of available elements and attributes for each set of best practices can be found in the appendix of each set of guidelines.

All CDL TEI best practices also assume that the electronic text being encoded is being derived principally, if not wholly, from an existing paper source document. That is, these guidelines are not expressly intended for projects creating born-digital texts, although they may be adapted for such use. These guidelines are intended for projects that are producing semi-diplomatic transcriptions of a source document with few if any editorial changes. While projects may choose not to reproduce the *look* or layout of a source document through their encoding, no emendation (meaning deliberate editorial change) of any textual element in the source document is permitted unless the project's emendation policy, spelling out what has been changed and what preserved, can be consistently applied and clearly explained in the document header.

The *CDL TEI Lite Best Practice Guidelines for Encoding Basic Texts* are the simplest set of TEI-based encoding guidelines provided by the CDL. They generally reflect the core practices behind all the more specific encoding models supported by CDL for different text types. Because the CDL TEI Lite guidelines are basic, if at all possible, projects should follow one of those more specific type guidelines. However, if the texts being encoded are very simple or the collection only requires minimal functionality, then the CDL TEI Lite guidelines and DTD may be used. All CDL TEI 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 [<http://www.tei-c.org/P4X/>] should be consulted.

Using These Guidelines

These guidelines are prescriptive. However, not all individual practices mentioned here are absolutely required for compliance to the standard. The following list provides the words and phrases that should serve as cues throughout this document as to whether a practice is required, recommended, or optional:

- **REQUIRED**

must, must not, will, will not, do, do not

Unless the practice is followed, the document will not be considered valid as a CDL TEI document. Where possible, these practices will be enforced by the DTD or schema.

- **RECOMMENDED**

should, should not

The recommendation should be followed if possible; it should only be violated if the encoder has a good reason for doing so. Where possible, these recommendations will be enforced by the CDL using a Schematron assertion language schema.

- **OPTIONAL**

may, may not, can, cannot

Although suggested, the practice is optional. Encoders may choose other valid strategies as necessary.

If a question arises that cannot be resolved through consulting these guidelines, the encoder should consult official TEI P4 [<http://www.tei-c.org/P4X/>] documentation. Throughout these guidelines, relevant sections of TEI P4 will be referenced using the following notation:

[P4: 11.2]

Chapter 1. General Instructions

1.1. File Management and ARKs

Every digital object submitted to the CDL, including objects that are associated files referenced by the main XML document, must be assigned an Archival Resource Key (ARK) that will serve as the object's unique and persistent identifier. Projects may obtain ARKs through the CDL for use in their encoding, or their files may automatically be assigned ARKs by the CDL upon ingest. The method by which a project's files will receive ARKs should be negotiated in advance and laid out in each project's submission agreement with the CDL.

For TEI files, each text's ARK will also be assigned as the value of the **id** attribute in the root element of the text's XML file. It will also be recorded in an <idno> element in the text's TEI header.

1.1.1. Naming

It is highly recommended that where possible the ARK also be used for naming TEI files, using the following convention:

ARK.xml, where "ARK" is the unique key assigned.

To facilitate the ingest of files, projects should use the following naming conventions for images, PDFs, and other associated content:

ARK_NAME.EXTENSION, where "ARK" is the unique key assigned, "NAME" is the result of whatever local naming convention has been applied to individual files, and "EXTENSION" is the normal file format extension (".gif", ".jpg", ".pdf", etc.).

type of file	ARK	file name
TEI	kt167nb66r	kt167nb66r.xml
GIF	ARK kt167nb66r	kt167nb66r_fig002.gif

1.1.2. Associated Content Files

All digital objects referenced as external entities by a TEI document must first be declared as entities at the beginning of the document. The entity declaration must give the object's entity reference and then define the reference using the object's system identifier. The system identifier must either be a system path relative to the document or, preferably, a URL. Ideally, to facilitate the preview and ingest of TEI objects, projects should make their documents and all associated content files (DTDS, images, pdfs, etc.) available via the web. Entity declarations must use the full object filename and the appropriate file format notation (e.g., GIF, JPG, or PDF).

```
<!ENTITY fig002 SYSTEM "http://www.server.domain/figures/kt167nb66r_fig002.gif" NDATA GIF
...
<figure id="fig002" entity="fig002" rend="block">
```

1.1.3. Image Files

The CDL will accept image files in either the GIF or JPEG format. If possible, two derivative images should be created for

each plate, figure, graphic, or other pictorial element that appears as a discrete element in the text. One of these derivatives should be at web resolution (72 ppi) and the same size as the figure in the printed text. The other image should be at higher resolution (300 ppi), again at original size, but not exceeding 768 pixels in width. In-line images, such as images of formulas, need only be provided in the low-resolution version. When necessary, images should be cropped and flipped for proper orientation for web display. For more information about the CDL's digital image standards, see the *California Digital Library Digital Object Standard: Metadata, Content and Encoding* [<http://www.cdlib.org/news/pdf/CDLObjectStd-2001.pdf>] and the *California Digital Library Digital Image Format Standards* [<http://www.cdlib.org/news/pdf/CDLImageStd-2001.pdf>].

The master version of the image (usually a TIFF) does not need to be submitted to CDL. However, projects interested in preserving master images for future use should consider submitting them to the UC Libraries Digital Preservation Repository [<http://www.cdlib.org/inside/projects/preservation/dpr/>], scheduled to launch in 2005.

If images are to be supplied in multiple resolutions, it will be necessary to encode this fact in a metadata record conforming to the Metadata Encoding and Transmission Standard (METS) schema.

```
<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 the CDL ingest team before constructing a METS record for objects with multiple resolutions.

1.2. Invoking the CDL TEI Lite DTD

All documents complying to these guidelines must explicitly invoke the CDL TEI Lite DTD. To do this, declare the the TEI XML DTD and include the prose, figures, and linking tag sets. Then include the CDL user extension files and the entity "CDL.lite". Other external entity declarations should directly follow. (See the section on associated files for instructions on how to declare entities.)

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

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

. . .
<!ENTITY fig002 SYSTEM "http://www.server.domain/figures/kt167nb66r_fig002.gif" NDATA GI
. . .
]>

```

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

1.3. Case Sensitivity

Please take note that XML is case-sensitive. All elements and attributes must be in the proper case to be valid. In the CDL TEI DTDs, all elements made up of compound words use the "camel case" format: e.g., "teiHeader" instead of "teiheader" or "TEIHEADER".

1.4. Character Encoding

Special characters in the text must be encoded using the Unicode Standard [<http://www.unicode.org/standard/standard.html>] (UTF-8) and documents must include "UTF-8" as the value of the **encoding** attribute in the XML declaration.

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

Special characters may be incorporated into a document directly as native Unicode (à) or may be represented by numeric character entities. These numeric character entities can take either the decimal (à) or hexadecimal forms (à). Characters must *not* be represented using named character entities (`), with the exception of those specifically exempted in the XML 1.0 Specification [<http://www.w3.org/TR/REC-xml/>]. These *must* be used to avoid validation errors:

character	description	Unicode
<	less than	<
>	greater than	>
&	ampersand	&

```
<p>The &lt;body> element contains the main body of the text.</p>
```

Unicode named character entities must also be used within attribute values that need to contain single or double quotation marks or apostrophes. Use the following named character entities to avoid a parser error:

character	description	Unicode
"	quotation marks	"
'	apostrophe or single quotation mark	'

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

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

1.5. Hyphenation

When encoding the text, take care not to transcribe end-line hyphens that have been introduced into the text as a result of typesetting. Record all hyphens that are required by the source for the correct spelling of a compound word or phrase. Similarly, record all hyphens that are absolutely necessary to the meaning of an expression, e.g., hyphens in dates, formulas, code, etc.

1.6. Metadata Encoding and Transmission Standard (METS) Record

The principal container for metadata at the CDL is a digital object's METS record. TEI documents should be submitted with as complete a METS record as possible. The CDL may generate METS records for projects that are unable to provide them. For more information, see The CDL METS Repository [<http://www.cdlib.org/inside/diglib/repository/>]'s web stie.

Chapter 2. Encoding Practice

2.1. Root Element

2.1.1. <TEI.2>

Each document should contain one and only one <TEI.2> root element. The **id** attribute is required and must contain the unique ARK assigned to the text in question.

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

2.2. Document Header

2.2.1. <teiHeader>

Generally, the <teiHeader> for each document must conform to the practices described in detail in the *California Digital Library Best Practice Guidelines for Encoding TEI Headers*. [../metadata] Those guidelines cover both mandatory practices as well as suggested or optional practices. It is often sufficient to follow the instructions there for encoding the mandatory minimal header. However, projects that will depend on the TEI header as their principal source of metadata (e.g., projects not providing their own METS records) are advised to use the recommendations for full header encoding.

CDL search indexing and metadata collection depend on using a crosswalk that maps individual TEI header elements to their Dublin Core Metadata Initiative [<http://dublincore.org/>] (DC) equivalents. A detailed list of which elements in the TEI header map to which elements in DC can be found in Appendices A and B of the CDL TEI Header guidelines [../metadata].

It is particularly important to note that every TEI document must make use of the <idno> element in the TEI header to record both the text's ARK and its local object identifier. Each must be given as the content of a separate <idno> element. The **type** attribute must be used to identify whether an "ARK" or "LOCAL" identifier is being given. These <idno> elements are essential to maintaining the link between the document and its various identities.

The following is an example of a minimal TEI header suitable for CDL TEI Lite documents.

```
<teiHeader type="CDL-TEI:LT">
  <fileDesc>
    <titleStmt>
      <title>[MF's personal journal on "Camouflage"] : electronic version</title>
      <respStmt>
        <resp>Text encoder:</resp>
        <name reg="Pollock, Alvin">Alvin Pollock</name>
      </respStmt>
    </titleStmt>
    <extent>ca. 3 Kb</extent>
    <publicationStmt>
      <publisher>The Bancroft Library</publisher>
      <pubPlace>UC Berkeley, Berkeley CA. 94720-6000</pubPlace>
      <date value="2001-00-00">2001</date>
      <idno type="ARK">ark:/13030/ft0b69n5bz</idno>
```

```

    <idno type="LOCAL">brk00002693_7a</idno>
  </publicationStmt>
  <seriesStmt>
    <title>Japanese American Relocation Digital Archives</title>
    <idno type="OTHER">http://jarda.cdlib.org/</idno>
  </seriesStmt>
  <sourceDesc>
    <biblFull>
      <titleStmt>
        <title>[MF's personal journal on "Camouflage"]</title>
      </titleStmt>
      <extent>1 p.</extent>
      <publicationStmt>
        <date>March 16, 1943</date>
      </publicationStmt>
    </biblFull>
  </sourceDesc>
</fileDesc>
</teiHeader>

```

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

2.2.2. <idno>

It is particularly important that each document makes use of the <idno> element in the TEI header to record both the text's ARK and its local object identifier. Each must be given as the contents of a separate <idno> element. The **type** attribute must be used to identify whether an "ARK" or "LOCAL" identifier is being given. These <idno> elements are essential to maintaining the link between the document, its persistent identifier, and its local identifier.

```

    <idno type="ARK">ark:/13030/ft0b69n5bz</idno>
    <idno type="LOCAL">brk00002693_7a</idno>

```

2.3. Text Structure

2.3.1. <text>

The <teiHeader> is directly followed by the mandatory <text> element, which fully contains the content of the source document being encoded. The <text> element contains three subelements, <front> for front matter (e.g., title pages, prefaces, and introductions), <body> for the main body of the text, and <back> for back matter (e.g., endnotes and appendices). Of these three, only <body> is required.

```

<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>]

2.4. Front Matter

The <front> element is used to contain the various components that make up front matter, including prefaces, introductions, and title pages. Each of these sections is normally contained within another structural element such as <titlePage> or a <divn>.

2.4.1. <titlePage>

Do not encode a <titlePage> unless the source document itself contains a formal title page. Title pages may use a number of formatting peculiarities, such as specific alignment, fonts, incidental images, etc. It is not necessary to attempt to reproduce the look of the title page in the document 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 the information 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>

```

2.4.1.1. <docTitle>, <titlePart>

The <docTitle> element is required within <titlePage>. Use <titlePart> within <docTitle> to encode individual formal titles, subtitles, and other subsidiary title parts as they appear on the title page. If there is more than one <titlePart> given, projects must use the **type** attribute to classify the various <titlePart>s. Supported **type** attribute values are "main," "subtitle," "alternate," and "abbreviated." Any <titlePart> without a **type** attribute will be considered and formatted as a "main" title. If there is more than one <titlePart>, then give the **type** attribute is mandatory for all of them.

```

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

```

2.4.1.2. <docAuthor>

Record here the names of authors and others responsible for the intellectual content of the document as they appear on the title page. Each <docAuthor> element will be displayed by the stylesheet on a single line. Therefore, projects 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 the names should be displayed on separate lines.

Projects may use the <name> element to surround each author's name. This practice is optional, but is particularly useful when more than one name has been encoded in a single <docAuthor>. The <name> element also allows projects to regularize names using the **reg** attribute.

In the content of <docAuthor> and <name>, names should be recorded as they appear 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. Again, the **reg** attribute may be used to correlate a name to an authority.

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

OR:

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

2.4.1.3. <byline>

Authors are frequently listed on the title page accompanied by a more explicit description of their role in the creation of the document; e.g., "foreword by" or simply "by." In such cases, encode both the <docAuthor>s and their statements of responsibility inside an encompassing <byline> element.

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

2.4.1.4. <docImprint>, <pubPlace> , <publisher>

Record the remaining publication information in <docImprint>. Within <docImprint>, use <pubPlace> and <publisher> in any order and as often as necessary to record every place of publication and every publisher respectively.

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

2.4.1.5. <docDate>

Record copyright and publication dates within <docDate> in <docImprint>. Do not include any associated text or symbols such as the word "copyright" or the symbol "©". Such words and symbols may be kept in the surrounding <docImprint> element. A regularized form of the date may be encoded in ISO 8601:2000 5.2.1.1 standard form (e.g., YYYY-MM-DD) in the **value** attribute of the <docDate> element. This is useful if document dates need to be consistently indexed.

```
<docImprint>New York Publishing Company &#xA9;<docDate value="1971.00.00"> 1971.</
```

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

2.5. Document Body

2.5.1. <body>

Containing the main body of the text, the mandatory <body> element is further subdivided into a hierarchy of nested divisions beginning with a mandatory <div1>. Use the **type** attribute in each <divn> to describe the type of section being encoded.

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

2.6. Back Matter

2.6.1. <back>

The optional <back> element may contain any number of <divn> elements containing advertisements, afterwords, indexes, bibliographies, appendices, or other sections that appear at the end of the document after the main body of the text. Use the **type** attribute in each <divn> to describe the type of back matter being encoded.

```
<back>
  <div1 type="appendix">
    <head>Photographs</head>
    <p>The author was a prolific photographer who. . .
    </p>
  </div1>
```

2.7. Divisions

2.7.1. <divn>

The <front>, <body>, and <back> elements in the document must use a hierarchical structure of numbered <divn> elements to identify their significant divisions. The elements <body> and <back> are both required to contain at least one <div1>. No unnumbered <div> or <div0> elements are permitted.

Each <divn> element throughout the text must have a unique **id** attribute to serve as an identifier. If necessary these can be added automatically on ingest by the CDL, depending on the project's submission agreement with the CDL.

All <divn>s must also contain a **type** attribute describing the kind of division being encoded. Every attempt should be made to supply the most specific and consistent **type** values possible for <divn> elements.

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

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

2.8. Division Headings

2.8.1. <head>

The <head> element is used to record division headings, such as chapter or section titles, and is used by the system for allow users to navigate easily from one section to another.

Specific guidelines are supplied below regarding where <head>s may or may not appear. Generally, record headings as they appear in the source document.

Headings may be supplied by the encoder if they are not available in the text but are necessary in order to provide a way of navigating to a particular division. Headings may also be supplied in cases in which a <head> is necessary to conform to rules about when they must appear.

Supplied headings should be enclosed in square brackets or signalled by some other convention expressly detailed in the <editorialDecl> of the <teiHeader>.

Title transcribed from text:

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

Title supplied by encoder:

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

It is good practice to provide a `<head>` tag for all major textual divisions. In any case, the following rules must be strictly followed:

1. If any `<divn>` at any level contains a `<head>`, then all of its sibling `<divn>`s at the same level must also contain a `<head>`. Therefore, if any `<div1>` uses a head, all `<div1>`s in the text must do so. If any `<div2>` contains a `<head>`, all other `<div2>`s nested with that `<div2>` in its parent `<div1>` must also contain `<head>`s, etc.
2. If a `<divn>` at any level is left without a `<head>`, then any subordinate `<divn>`s below the headless `<divn>` are not permitted to have `<head>`s. Conversely, if any subordinate `<divn>` contains a head, the parent `<divn>` must also contain a `<head>`.

The following example is incorrect because one of the `<divn>` descendants contains a `<head>` but none of its ancestors contain one. If the rules are strictly followed, the single `<div4>` with a `<head>` forces all other `<div>`s in the tree to contain `<head>`s:

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

Multiple `<head>` elements may be differentiated using the **type** attribute (e.g., "subtitle" for a subtitle).

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

2.9. Paragraphs

2.9.1. `<p>`

The paragraph is the fundamental organizational unit for all prose texts. Paragraphs are encoded within `<p>`s, which, by default, begin a new line and are displayed with the first line indented. To dictate a different display, use the **rend** attribute in `<p>`. Please see the section on alignment and indentation for a list of available **rend** values.

```
<p>In another moment down went Alice after it, never once
considering how in the world she was to get out again.</p>
```

[P4: 6.1] [<http://www.tei-c.org/P4X/CO.html#COPA>]

2.10. Page Breaks and Milestones

Milestones are empty elements (<lb>, <milestone>, <pb>) that serve a function in the text analogous to the one mileposts serve on a road. They are used to mark significant points in the text, often beginnings or endings of sections, that exist outside the hierarchy of <divn> containers.

2.10.1. <pb>

Projects must use the empty <pb> element to mark the beginning of each physical page of the source document (including the first page). The <pb> element should be placed at the beginning of each page, but entirely within any overlapping <divn>. Never encode <pb>s between <divn> elements. All such interstitial page breaks should be encoded as if they belonged to the nearest subsequent <divn>, before the <head> element. If a page break occurs in the middle of a smaller block element (e.g., <p>), it can simply be encoded there.

If desired, the **n** attribute of <pb> may be used to record page numbers as they appear in the source document so that the system can subsequently render those page numbers for display. Do not supply page numbers if they do not exist in the source document. Page numbers should be recorded using the **n** attribute of the <pb> element at the beginning of the page, regardless of where the number appears on the document.

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

If a page number is given, the **id** attribute is also highly recommended. If anything is linked to the page breaks (such as an index entry or table of contents that refers to pages), the **id** attribute is required.

```
<p>of the Sea, <ref target="p1" type="pageref">1</ref></p>
. . .
<pb n=1 id="p1"/>
```

2.10.2. <lb>

The <lb> element marks the start of a new line. Use this element only when it is absolutely essential to preserve line breaks as they appear in the source document. (Note that the <lb> tag is intended for producing line breaks in prose only. the <l> element must be used to encode lines of verse.)

```
<p>When I approached the door, I saw that it's knocker yawned as a great
<lb/>0
<lb/>before me, impossibly heavy. . . </p>
```

2.10.3. <milestone>

The empty <milestone> element may be used to mark significant boundaries between sections of text that are neither page breaks nor normal divisions. For instance, it may be used to encode the decorative section breaks common to monographs. The **unit** attribute is required to describe the kind of break being marked. The **n** attribute must be used to record any characters or symbols that are used to create the boundary.

```
<milestone unit="endPart" n="&2766;" />
```

```
<milestone unit="endPart" n="****" />
```

[P4: 6.9.3] [<http://www.tei-c.org/P4X/CO.html#CORS5>]

2.11. Typographical Phenomena and Formatting

2.11.1. <hi>

Record font changes and other typographical highlighting with the <hi> element. Use the required **rend** attribute to record the type of font shift employed in the source document. Unless otherwise stated in the <editorialDecl>, the value of the **rend** attribute must convey and ultimately display (if possible) the actual marking in the source document. In other words, do not use <hi> to introduce editorial changes to a text's typesetting.

When text with special formatting has already been tagged for other structure or content, and when the special formatting is consistent, the **rend** value can be applied directly to the encompassing tag. For example, if the contents of <name> are underscored, or if the contents of <p> are entirely in bold font, then the **rend** values of those tags can be defined accordingly. Because **rend** is a global attribute, it is available for all TEI elements. When special formatting does not coincide perfectly with an encompassing tag (as is often the case), <hi> is used to surround the special text.

```
<p><hi rend="underline">Where</hi> did he go?</p>
```

```
<head rend="smallcaps">The Last Stand</head>
```

The CDL supports the following **rend** values for display:

value	display
normal	standard font for the document; unemphasized, unhighlighted text; should be used to format unemphasized text in the middle of an emphasized passage
mono	mono-spaced font, e.g., Courier
italic	italics
smallcaps	small caps
bold	bold
bolder	extra bold
lighter	extra light
underline	underscored

value	display
overline	written with a line drawn above the text
strikethrough	strikethrough
subscript	below the baseline of standard text
superscript	above the baseline of standard text
hide	do not display

Projects requiring more specialized display may include syntax from the Cascading Style Sheet (CSS) [<http://www.w3.org/Style/CSS/>] standard in the **rend** attribute.

```
<p rend ="color: white; background-color: red">This text will be white on a red background</p>
```

2.11.2. Nested <hi> Tags

When multiple **rend** values are required for a single element, repeat <hi> elements as necessary. For instance, in the following example, the word "wow" is rendered in both bold and italics as "**wow**".

```
<hi rend="bold"><hi rend="italic">wow!</hi></hi>
```

Remember that once a **rend** value has been applied to a tag, the display is applied to the entire contents of that tag unless it is explicitly negated by another tag. For instance, the tagging

```
<hi rend="bold">w<hi rend="italic">ow!</hi></hi>
```

will produce the word "**wow**".

On the other hand, the tagging

```
<hi rend="bold">w<hi rend="normal"><hi rend="italic">ow!</hi></hi></hi>
```

will produce the word "**w ow**".

2.11.3. <emph>

If desired, the <emph> element may be used instead of <hi> to mark a typographic shift that explicitly conveys emphasis rather than simply a change in typography or other meaning. In the following example, the word "very" is underscored to provide emphasis.

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

The same **rend** values available for `<hi>` are also available for `<emph>`, as they are for all **rend** attributes in any element.

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

2.11.4. Alignment and Indention

Alignment and indention of text can also be represented using the **rend** attribute in `<hi>` or any other encompassing tag. Available **rend** attribute values for alignment are:

value	display
left	justify left, ragged right, initial indent
center	center
right	justify right, ragged left, initial indent
justify	fully justify, initial indent
indent	standard paragraph indent
hang	hanging indent
blockindent	full block indent
blockquote	full block indent used for quotes (<code><quote></code>)
noindent	no initial indent

Projects requiring more precise alignment of text may also use CSS [<http://www.w3.org/Style/CSS/>] language within the **rend** attribute to describe the alignment required.

```
[4em hanging indent]
<p rend="text-indent: -4em; margin-left: 4em">
```

Note that all `<p>`s are flush left with an initial indent by default, so any paragraph that should not be indented must be given a **rend** value of "noindent".

2.12. Language Shifts

2.12.1. `<foreign>`

Use the `<foreign>` element to tag text that appears in a language that will require the use of a different character set or writing direction. The **lang** attribute must contain the name of the applicable language as given in the `<language>` element [/stwg/metadata/META_BPG.html#d0e294] of the TEI header. Note that the language *must* be declared in the TEI header in order for this attribute to function. The enclosed text should be input using the appropriate Unicode character entities. (See the section on character encoding.)

```

    <profileDesc>
    <langUsage>
      <language id="Greek">(Range: 0370-03FF)</language>
    </langUsage>
    . . .
      <foreign lang="Greek">&#0371;&#0372;&#0399;</foreign>

```

```
###
```

2.13. Quotations

Quotations that are set apart from the rest of the text by quotation marks need not be specially encoded. Quotation marks are normally left intact in the text and, if possible, recorded in the form that they appear (i.e., straight or curly, single or double). (The exceptions to this rule are quotation marks around <title>s in bibliographic citations that use the **level** attribute to provide their formatting [see the section on bibliographies].)

Quotations that employ formatting beyond the simple use of quotation marks must be specifically tagged. Simple block quotes containing only one paragraph may be recorded using <p>.

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

2.13.1. <quote>

Quotes comprising multiple paragraphs or lines of verse should be enclosed in the <quote> element, with individual paragraphs contained in <p>s and lines of verse contained in <lg> and <l>.

```

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

<quote>
  <lg>
    <l>What you have seen to love in me</l>
    <l>I do not know.</l>
    <l>What I have seen to love in thee</l>
    <l>No word can show. </l>
    <l>But word or knowledge, dear, we lay aside.</l>
    <l>We need them not for compass or for guide.</l>
    <l>By love we go.</l>
  </lg>
</quote>

```

2.13.2. <cit>

If desired, quotations that are accompanied by citations may be encoded using <cit>. Enclose both the quote and the citation within <cit>. The text of the quote should be further enclosed within <q> and <p> as necessary, and the bibliographic citation should be further enclosed within <bibl>.

```
<cit>
  <quote>
    <l>Since I can do no good because a woman</l>
    <l>Reach constantly at something that is near it.</l>
  </quote>
  <bibl>
    <title>The Maid's Tragedy</title>
    <author>Beaumont and Fletcher</author>
  </bibl>
</cit>

<cit>
  <quote>
    <lg>
      <l>`Twas brillig, and the slithy toves</l>
      <l>Did gyre and gimble in the wabe:</l>
      <l>All mimsy were the borogoves,</l>
      <l>And the mome raths outgrabe.</l>
    </lg>
  </quote>
  <bibl>"Jabberwocky"--Lewis Carroll</bibl>
</cit>
```

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

2.14. Verse

2.14.1. <divn> in Verse

Generally, verse or verse fragments in a text should be enclosed within a separate <divn> element with an identifying **type** attribute. Projects must enclose a poem in a <divn> if they wish to attach a searchable, indexable title to the poem using <head> or if they wish to encode a <closer> at the end of the poem. The most common **type** attribute values for verse are:

```
verse
poem
sonnet
drama
free-verse
song
```

If projects do not wish to enclose a poem within a separate <divn>, they may simply enclose its lines using the mandatory <lg> element. (See below.)

2.14.2. <head> in Verse

Projects may use the <head> element for all titles, subtitles, etc., for verse encoded within a <divn>, bearing in mind the rules for using <head> within <divn>. When more than one <head> is required, use the **type** attribute to describe the different type of headings or titles being applied. Any <head> element for verse that does not have a **type** attribute will be

considered a "main" title.

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

2.14.3. <l>

Individual lines of verse must be surrounded by the <l> tag. Lines that are numbered may use the **n** attribute to encode the line number. Use the **rend** attribute as necessary to provide proper indentation.

```
<l n="5" rend="indent">
```

2.14.4. <lg>

Regardless of whether verse is contained within its own <divn>, groups of lines must be encoded within the <lg> element, with each individual line also encoded in the <l> element. The <lg> tag is used to identify groups of lines that carry coherent poetic structure (i.e., function as a formal unit, such as a stanza) within a poem. The type of structure may be identified with the **type** attribute. Some available *type* values are:

```
stanza
verse
paragraph
couplet
quatrain
fragment
refrain
```

The value "fragment" should be used for line groups that do not carry poetic structure.

```
<div1 type="poem">
  <lg type="stanza">
    <l>How doth the little crocodile</l>
    <l>Improve his shining tail,</l>
    <l>And pour the waters of the Nile</l>
    <l>On every golden scale!</l>
  </lg>
</div1>
```

The following text could be tagged in different ways:

```
`Repeat, "You are Old, Father William," said the Caterpillar.
  Alice folded her hands, and began:--
  `You are old, Father William,' the young man said,
    `And your hair has become very white;
```

And yet you incessantly stand on your head-
Do you think, at your age, it is right?'

within `<divn>`:

```
<div1 type="chap5">
. . .
<p>`Repeat, "You are Old, Father William,"' said the Caterpillar.</p>
<p>Alice folded her hands, and began:--</p>
<div2 type="poem">
<head type="poem-title" rend="center">[You Are Old, Father William]</head>
<lg type="stanza" rend="blockindent">
<l>`You are old, Father William,' the young man said,</l>
<l rend="indent">`And your hair has become very white;</l>
<l>And yet you incessantly stand on your head-</l>
<l rend="indent">Do you think, at your age, it is right?'</l>
</lg>
</div2>
```

or without `<divn>`:

```
<div1 type="chap5">
. . .
<p>`Repeat, "You are Old, Father William,"' said the Caterpillar.</p>
<p>Alice folded her hands, and began:--</p>
<lg type="stanza" rend="blockindent">
<l>`You are old, Father William,' the young man said,</l>
<l rend="indent">`And your hair has become very white;</l>
<l>And yet you incessantly stand on your head-</l>
<l rend="indent">Do you think, at your age, it is right?'</l>
</lg>
```

2.15. Notes

2.15.1. `<note>`

Use the `<note>` element to encode notes, using the **place** attribute to indicate the location of the note. Available **place** values are:

value	type of note
end	endnote, note appears at the end of a chapter, part, or volume
foot	footnote, note appears at the foot of the page
inline	note appears as a marked section in the body of the text

Notes without the **place** attribute will be considered in-line. For notes that are tagged at the point of reference, the

numbers attached to the notes (as distinct from reference numbers that are located elsewhere) are normally recorded as the value of the **n** attribute and should not be included in the text of the note itself. Similarly, dingbats, crosses, daggers, and the like used to label notes for referencing may also be recorded as Unicode characters within the **n** attribute. A separate `<ref>` is not necessary. If a note is targeted by a `<ref>` elsewhere, it must contain a unique **id** attribute. Be sure to enclose the contents of notes in `<p>s` or other appropriate block elements if necessary. (See the section on internal linking for more information about `<ref>s`.)

2.15.2. In-line notes

In-line notes may be tagged directly in place.

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

2.15.3. Footnotes

Footnotes (those references, notes, and citations appearing at the bottom of the page) must be encoded where they are referenced. In other words, at the location of the footnote reference in the text, embed the `<note>` itself in place. If a footnote is tagged in place *and* the **n** attribute contains the note's reference number, projects must not encode a separate `<ref>` with that same number in the same location. The result would be two duplicate numbers appearing in place at the point of reference. However, if no **n** attribute is given in `<note>`, then a separate `<ref>` may be used in place. In either case, other references to that footnote from other locations in the text may be tagged with `<ref>`. If a footnote is targeted by any `<ref>` anywhere in the text, it must include an **id** attribute. (See the section on internal linking.)

```
<p>...Whites, however, did not vote to transfer power
<hi rend="italic">to</hi> the black majority, as the
media reported, but only to share power.
<note id="fn0.1" place="foot" n="*">
<p>The use of racial and ethnic labels is not meant
to reproduce, uncritically...</p></note>...</p>
```

2.15.4. Endnotes

Endnotes (those appearing at the end of a chapter, section, or other significant textual division) must be encoded where they appear in the document, in a separate `<divn>` if necessary. For an endnote to function properly, the reference to the note in the text must be tagged with `<ref>` and each endnote `<note>` must carry an **id** attribute. Further, if projects wish to allow users to link directly from the note back to its reference in the text, then the **id** and **corresp** attributes must also be properly used in `<ref>` and `<note>` respectively. (See the section on internal linking for more detailed instructions.)

```
<p>...falsely assumed South Africa to be the only developed
capitalist country "[that] is not only 'objectively' ripe for
revolution but has actually entered a stage of overt and
seemingly irreversible revolutionary struggle."
<ref target="bn0.1" id="d0e912" type="noteref">1</ref> ...</p>
```

....

```
<div2 id="d0e1020" type="endnotes">
  <head type="main">Notes</head>
  <note id="bn0.1" place="end" n="1" corresp="d0e912">
    <p>Paul M. Sweezy and Harry Magdoff, "The Stakes in
      South Africa," <hi rend="italic">Monthly Review,
    </hi> April 1986.</p>
  </note>
</div2>
```

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

2.16. Lists

2.16.1. <list>

Individual items in a list must be encoded as <item>s within <list> rather than as a series of <p>s or <l>s. Use the <list> element's **type** attribute to define the type of list appearing in the document. Valid **type** attributes are:

value	type of list
ordered	lists with sequential markers
bulleted	marked or bulleted lists
simple	unmarked or unnumbered lists
gloss	definition lists (e.g., glossary, chronology, etc.) consisting of a term encoded in <label> and a definition or expansion of the term encoded in <item>
ordered	numbered lists
label	non-gloss lists whose items are each labeled with a <label>

Nest lists as appropriate, noting that they will be automatically indented to reflect the nesting. Use the <head> element to provide headings for lists.

2.16.2. Standard Ordered Lists

Encode lists that include sequential markers, numbers, or letters as <list type="ordered">. Use the **rend** attribute to describe the kind of sequential system used. Each item in the list is encoded as an <item>, *without* the sequential marker. The **rend** attribute will tell the stylesheet what kind of enumerative system to supply for display. If no system is specified in the **rend** attribute, then the default system of "arabic"--meaning arabic integers starting with "1."-- will be applied. The available **rend** values are as follows.

value	enumerators
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	non-standard enumerations encoded within each <item>'s n attribute (see below)

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>
```

2.16.3. Non-standard Ordered Lists

Lists that use a non-sequential or otherwise non-standard method of enumeration may still carry the **type** attribute value of "ordered" if the specific mark of numeration may be explicitly supplied in the **n** attribute of each individual `<item>` element. Whatever is encoded as the value of the **n** attribute will be exactly displayed as the enumerator for the item. Therefore, don't forget to include punctuation if it is desired. In such cases, set the `<list>`'s **rend** attribute to "supplied."

- 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 all `<item>`s in a `<list rend="supplied" type="ordered">` must contain an **n** attribute, even if some of the items conform to the standard enumerative conventions. Again, never encode the sequential marker within the text of the `<item>`

as well. Such encoding will usually result in two duplicate markers appearing before each <item> in the list. **n** attribute.

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

2.16.4. <label>

Rather than enumerators, items in a <list type="gloss"> have labels, such as headwords in a glossary or dates in a chronology. The <label> element is used to capture each label immediately preceding its associated <item>.

```
<list type="gloss" rend="label">
<label>1835</label><item>born in Florida, MO</item>
<label>1848</label><item>apprenticed</item>
```

2.17. Bibliographies

2.17.1. <bibl>, <listBibl>

Individual bibliographic citations should be encoded using the <bibl> element. Groups of <bibl>s are further contained within a <listBibl>.

The <bibl> element allows unstructured bibliographic data, including standard bibliographic elements as well as uncontained text such as more discursive or descriptive citations or annotation. Unlike the stricter bibliographic containers found in the TEI, <bibl> allows the encoder some latitude both in the order of subelements and the level of encoding.

There are no elements absolutely required within <bibl>. However, most projects will most likely take advantage of the following: <author>, <date>, <title>, <pubPlace>, <publisher>, and <biblScope>.

```
<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>
```

2.17.2. <title> levels

Projects using the <title> element may also use its **level** attribute to define the type of title being provided and dictate the standard typographic styling used to display the title. Therefore, <title>s that carry a **level** attribute do not need to be tagged again for italics, quotation marks, and the like. Titles that require special formatting not supported by the available **levels** can use the **rend** attribute to dictate the styling required. The supported attribute values and their resulting display are as follows:

value	type of title	type of styling
a	analytic title (article, poem, or other	surrounded in quotation marks

value	type of title	type of styling
	item published as part of a larger item)	
m	monographic title (book, collection, or other item published as a distinct item, including single volumes of multi-volume works)	italics
j	journal title	italics
s	series title	italics
u	title of unpublished material (including theses and dissertations unless published by a commercial press)	surrounded in quotation marks

2.17.3. <note> in Bibliographic Citations

The CDL TEI DTDs all allow the <note> element within <bibl>. Use it to record notes, including in-line bibliographic annotation and footnotes, that occur within bibliographic citations.

```
<bibl><title level="m">Alice's adventures in Wonderland</title> by
<author>Lewis Carroll</author>.
<pubPlace>London</pubPlace>:
<publisher>Macmillan</publisher>,
<date value="1869.00.00">1869</date>.
<note>This work is remarkable example of the intersection of mathematics and literature.
```

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

2.18. Internal Links and Cross References

Internal references and links can take many forms: numbers in the text that point to endnotes, page numbers in indexes that point to specific pages, pointers to specific sections of the text (e.g., "See Section 2A"), or short form bibliographic references (e.g., "Baxter 1978"). The practice described in this section applies only to references pointing to elements within the same file. See the section on external references to point to locations outside the document.

2.18.1. <ref>

Internal references will be encoded using the <ref> element and are required to have both a **target** and **type** attribute to indicate the **id** of the element being targeted and the nature of the target. (No specific system need be employed for creating **ids** in the elements being targeted as long as they are unique and begin with a letter character [e.g., id="id001"].) The following **type** attribute values are supported for <ref>:

value	type of reference
citeref	bibliographic citation reference
figref	figure reference
fnoteref	footnote reference
formularef	formula reference
noteref	endnote or general note reference

value	type of reference
pageref	reference to a <pb> element, such as would be used in an index
secref	section reference, usually used to refer to a chapter or subsection.
tableref	table reference

(Note that the use of a <ref> for footnotes is normally optional as the in-line presence of the <note> will automatically create a reference. References to the footnote from other locations are to be treated as <ref>s. See the section on footnotes for more information.)

```
<ref target="enote1" type="noteref">1</ref>
<note id="enote1" place="end" n="1">
. . .
</note>
```

In order to create a bidirectional link (i.e., from the reference [i.e., <ref>] to the referenced object [e.g., <note>] and then from the object back to the reference), projects must also include a unique **id** attribute in the <ref>. The value of the **id** in <ref> is then recorded in the **corresp** attribute of the element that is being referenced.

```
<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>]

2.19. External Objects

2.19.1. <xref>

Use the <xref> element to refer to objects or locations outside of the encoded document. There are six attributes available for <xref>. Take care to note which of these are required.

attribute	use	possible values	required?
doc	contains the object's entity name	[local entity name; must resolve to a valid declared entity]	required when href is not used
href	contains the external URI, may be URL or ARK	[external URI (e.g., URL or ARK)]	required when doc is not used
type	indicate the type of object being linked to	obj mets	required

attribute	use	possible values	required?
		url pdf sound video stream	
rend	defines the way the linking takes place	new replace embed none	required
from	contains the starting location of the portion of the digital object being linked to; also used to record single locations within objects	[usually a unique id on a structural element]	optional
to	contains the ending location of the portion of the digital object being linked to	[usually a unique id on a structural element; not required when only a single location in the object is being linked to]	optional

Note that every <xref> must have either a **doc** attribute or an **href** attribute or both.

The following table describes the actions dictated by the **rend** attribute:

value	resulting action
new	a new window displaying the referenced external object appears
replace	document view replaced by the referenced external object
embed	the referenced external object is embedded in place
none	no action

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 object.

PDF document:

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

Result: new window displaying a PDF of Chapter 19.

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

2.20. Graphic Elements

The encoding of graphic elements such as illustrations, formulas, and tables requires special care to preserve both the information contained therein and, as far as possible, the presentation.

2.20.1. <table>

Projects using the CDL TEI Lite best practice guidelines should use the default TEI table scheme as outlined in TEI P4 rather than the full XHTML table module required by the CDL guidelines for other text types. The TEI model is fairly limited but is much easier to use than the XHTML table module.

```
<table cols="3" id="tab001">
  <head>PERCENTAGES OF THE EARTH'S SURFACE</head>
  <row>
    <cell>Latitude</cell>
    <cell>%</cell>
    <cell>Cumulative %</cell>
  </row>
  <row>
    <cell>40 N 30 W</cell>
    <cell>8.68</cell>
    <cell>8.68</cell>
  </row>
```

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

2.20.2. <figure>

Figures, charts, plates, formulas, or any other component of the text that must be delivered as an image must be encoded using the <figure> element. Any <figure> must contain a unique **id** attribute and an **entity** attribute that contains a valid entity name that resolves to a real file. The entity named in the **entity** attribute must be declared at the beginning of the document in order for the document to validate and function properly during ingest and preview. See the sections on associated files and image files for detailed instructions on how to create entities and produce image files. The **rend** attribute is also required. The following **rend** values are available:

value	display
inline	in-line as part of a text string
block	as a block separate from the surrounding text
popup	linked to a higher resolution version; for pop-up figures use

value	display
	the following syntax: <code>rend="popup(ENTITY_NAME)"</code> , where the value in the parentheses is a valid entity name

Figure captions may be encoded in the `<head>` element within `<figure>` using the the **type** attribute value "caption".

```
<!ENTITY fig001 SYSTEM "http://www.server.domain/figures/fig001.gif" NDATA GIF>
<!ENTITY fig001_h SYSTEM "http://www.server.domain/figures/fig001_h.gif" NDATA GIF>
]>

<figure id="fig001" entity="fig001" rend="popup(fig001_h)">
  <head type="caption">Bottom topography in the South Atlantic Ocean.</head>
</figure>
```

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

2.20.3. Formulas

2.20.3.1. Formulas in `<figure>`

The difficulty of encoding mathematical and chemical formulas almost always makes it necessary for projects to submit an image of a formula rather than a marked-up representation. To provide the image of a formula, use `<figure>`.

```
<!ENTITY formula001 SYSTEM "http://www.server.domain/kt168nb88r_formula001.gif" NDATA GIF>
<!ENTITY fig001_h SYSTEM "http://www.server.domain/figures/formula001_h.gif" NDATA GIF>
]>

<figure id="formula001" entity="formula001" rend="inline">
```

2.20.3.2. `<formula>`

The CDL also supports the encoding of TeX formulas within the TEI's `<formula>` element. To encode TeX formulas, give the **notation** attribute a value of "TeX" and use the **rend** attribute to indicate whether the formula should be displayed "inline" or as a "block". Projects that wish to give both a TeX expression and an image of the formula may do both.

```
<formula notation="TeX" rend="block">
\l
\sigma_{s, \vartheta, p} = ({{\rho}_{s, \vartheta, p}} -1))1000.
\l
</formula>
```

[P4: 22.2] [<http://www.tei-c.org/P4X/FT.html#FTFOR>]

2.21. Arbitrary Containers and Segments

Arbitrary containers (<ab> and <seg>) can be nested virtually anywhere in the document and therefore can be used sparingly to resolve otherwise impossible encoding problems. When a necessary element is not valid in the location where it should logically go within a TEI document, an arbitrary container can be inserted in the correct place instead. The text can then either be tagged directly as the content of the arbitrary container, or it can be tagged first with the desired element, which is then dropped into the arbitrary container.

Arbitrary containers may also be used when no other available container element is appropriate for the text being marked up. This usage, however, should be very limited.

The **type** attribute is required for both <ab> and <seg> elements. Suggested attribute values for **type** are "figure", "illgrp", "tblgrp", and "text". Projects may assign other values as needed.

2.21.1. <seg>

Use <seg> to contain a segment of text or an element that may normally appear in a paragraph but needs to be encoded inside another element in which it is not otherwise allowed.

```
<address>
  <addrLine>The Compton Hotel<seg type=figure><figure id="seal1" entity="fig001"></addr
  <addrLine>1515 42nd Street</addrLine>
  <addrLine>Chicago, IL</addrLine>
</address>
```

2.21.2. <ab>

Use <ab> to contain element that may normally appear in a paragraph, but needs to be encoded directly into a major division such as a <divn> where it is not otherwise allowed.

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

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

Chapter 3. Quality Assurance

3.1. Validation

All documents must parse correctly before being submitted to the CDL. All texts will be validated on ingest and rejected if errors are detected.

3.2. Best Practice Checking

In addition to being validated against the supplied DTDs, documents will be checked for conformance to the appropriate CDL TEI best practice guidelines using a Schematron assertion language schema. Users can check their documents on their own by using the CDL Text Preview page

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

3.3. Proofreading

Proofreading the actual text of submitted documents is the responsibility of the contributor. It is highly recommended that all texts at least be spot-checked for major errors before submission. 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 TEI Lite documents and their attributes. Attribute value definitions take one of the following forms:

ENTITY	Entity Name defined in an entity declaration (<!ENTITY fig1 SYSTEM "fig1.gif" NDATA GIF>)
ID	Unique ID
ID REFERENCE	Reference to an existing ID
TEXT	Unrestricted text
URI	Uniform Resource Indicator
(OPTION1 OPTION2)	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
----	----	----------

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* .

<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:

id	ID	OPTIONAL
key	TEXT	OPTIONAL
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* .

<notesStmt>

Notes statement. Collects together any notes providing information about a text additional to that recorded in other parts of the bibliographic description.

Attributes:

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

See Also *Document Header* .

<opener>

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

Attributes:

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

See Also *Division Openers and Closers* .

<p>

Paragraph. Marks paragraphs in prose.

Attributes:

rend	(blockquote center hang indent	OPTIONAL
------	--------------------------------------	----------

noindent | left | right)

<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* .

<row>

Contains one row of a table.

Attributes:

id	ID	OPTIONAL
role	TEXT	OPTIONAL

See Also *Tables* .

<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.

Attributes:

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

See Also *Division Openers and Closers* .

<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: head, row, cell.

Attributes:

id	ID	OPTIONAL
cols	TEXT	OPTIONAL
rows	TEXT	OPTIONAL

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* .

<trailer>

Trailer. Contains a closing title or footer appearing at the end of a division of a text.

Attributes:

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

See Also *Division Openers and Closers* .

<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* .