

California Digital Library TEI Best Practice Guidelines for Encoding Manuscripts

California Digital Library Structured Text Working Group

California Digital Library TEI Best Practice Guidelines for Encoding Manuscripts

California Digital Library Structured Text Working Group
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
Using These Guidelines	viii
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 Manuscript DTD	2
1.3. Case Sensitivity	3
1.4. Character Encoding	3
1.5. Hyphenation	4
1.6. Extent of Encoding	4
1.7. 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
2.3. Text Structure	10
2.3.1. <text>	10
2.3.2. <group>	11
2.4. Front Matter	12
2.4.1. <front>	12
2.4.2. <titlePage>	12
2.5. Document Body	14
2.5.1. <body>	14
2.6. Back Matter	14
2.6.1. <back>	14
2.7. Divisions	15
2.7.1. <divn>	15
2.8. Division Headings, Openers, and Closers	16
2.8.1. <head>	16
2.8.2. <epigraph>	17
2.8.3. <opener>, <closer>	18
2.8.4. <byline>	19
2.8.5. <dateline>	20
2.9. Paragraphs	20
2.9.1. <p>	20
2.10. Page Breaks and Milestones	20
2.10.1. <pb>	21
2.10.2. <lb>	22
2.10.3. <milestone>	22
2.11. Typographical Phenomena and Formatting	23
2.11.1. <hi>	23
2.11.2. Nested <hi> Tags	24
2.11.3. <emph>	25
2.11.4. Alignment and Indention	25
2.12. Language Shifts	26
2.12.1. <foreign>	26
2.13. Changes in Hand	26
2.14. Quotations	27
2.14.1. <quote>	27
2.14.2. <cit>	28

2.15. Speech	28
2.15.1. <sp>	28
2.15.2. <speaker>	29
2.16. Verse	30
2.16.1. <divn> in Verse	30
2.16.2. <head> in Verse	30
2.16.3. <l>	30
2.16.4. <lg>	31
2.16.5. <closer>	32
2.17. Notes	32
2.17.1. <note>	32
2.17.2. In-line notes	34
2.17.3. Footnotes and Marginal Notes	34
2.17.4. Endnotes	35
2.18. Editorial Intervention	35
2.18.1. Editorial Notes	35
2.18.2. <sic>, <orig>	36
2.19. Gaps, Illegible Text, Damage	36
2.19.1. <gap>, <unclear>	36
2.20. Revisions in Manuscripts	37
2.20.1. <add>, 	37
2.20.2. Revisions in Another Hand	37
2.21. Names, Dates, and Addresses	37
2.21.1. <name>	37
2.21.2. <date>	38
2.21.3. <address>, <addrLine>	38
2.22. Lists	39
2.22.1. <list>	39
2.22.2. Standard Ordered Lists	39
2.22.3. Non-standard Ordered Lists	40
2.22.4. <label>	41
2.23. Bibliographies	41
2.23.1. <bibl>	41
2.23.2. <title> levels	42
2.23.3. <note> in Bibliographic Citations	42
2.24. Internal Links and Cross References	42
2.24.1. <ref>	42
2.25. External Objects	43
2.25.1. <xref>	43
2.26. Graphic Elements	45
2.26.1. Tables	45
2.26.2. <figure>	46
2.26.3. Formulas	46
2.27. Arbitrary Containers and Segments	47
2.27.1. <seg>	47
2.27.2. <ab>	47
3. Quality Assurance	49
3.1. Validation	49
3.2. Best Practice Checking	49
3.3. Proofreading	49

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 Best Practices for Manuscripts are for the encoding of manuscript material. For the purpose of these guidelines, manuscripts are defined as documents that bear physical evidence of human inscription or alteration of the text they contain, and that are valued in large part because of that text, as well as when, how, and by whom it was inscribed. Such documents include literary manuscripts; holograph, typewritten, or amanuensis letters; memos, notebooks, checks, photographs, business documents, or any physical document on which handwritten, typed, or carved text has been inscribed. In order to narrow the scope of this definition somewhat, certain kinds of documents have been excluded. The guidelines exclude published documents (e.g., books with marginalia), works created by medieval scribes, ancient texts, and dramatic works. The guidelines are purposely skewed toward capturing semantic content rather than physical details of a document. Therefore, projects seeking to provide full physical description of artifactual documents will need to supplement these guidelines. Still, the range of documents covered by these guidelines is large enough to make for some fairly sweeping generalizations, and that range limits the number of demands the guidelines can make. All projects digitizing manuscripts are encouraged to consult the full TEI P4 for more specific guidance on issues and phenomena they will no doubt encounter.

Projects that wish to emend their texts may choose to record specific instances of emendation within a critical apparatus. These guidelines offer no instructions on the use of TEI P4 for creating such an apparatus. Therefore, projects intending to emend manuscripts and/or produce a critical text from multiple documents need to consult TEI P4.

As with all CDL TEI guidelines, this document is meant to be used in conjunction with full documentation for TEI P4. Where an issue is not directly addressed in these guidelines, official 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 Manuscript DTD

All documents complying to these guidelines must explicitly invoke the CDL TEI Manuscript DTD. To do this, declare the the TEI XML DTD and include the prose, figures, transcription, and linking tag sets. Then include the CDL user extension files and the entity "CDL.ms". 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.transcr "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.ms "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&gt; 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.

Note that idiosyncratic spelling in manuscripts may mean that it is difficult to determine whether an end-line hyphen should be preserved in the transcription or not. Whatever method is used to resolve end-line hyphens should be recorded in the document's <editorialDecl> in the TEI header.

1.6. Extent of Encoding

Manuscripts transcribed following these guidelines should be transcribed and encoded in full such that all pages with authorial text and/or non-authorial commentary (including revisions and marginal notes by someone other than the original author) appear in the encoded transcription. Notes and other additions to the original manuscript for the purpose of sale or of cataloging (prices, catalog numbers, docketing by hands other than those who created, sent, or received the document) may be excluded if their exclusion is consistent and documented in the <editorialDecl> in the TEI header. Any part that is encoded but need not be displayed or accessed may be commented out of the XML file. A project's specific policies regarding what has been encoded and what left out, including policies adopted at the suggestion of these guidelines, must be articulated in the file's <editorialDecl> in the <teiHeader>.

1.7. 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

Generally, the <teiHeader> for each document must conform to the practices described in detail in the *CDL 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, as long as projects also meet additional requirements for manuscripts outlined below. 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. The instructions below include only specifications that modify or expand on the CDL TEI Header guidelines in some way. Therefore, they *cannot* be used alone.

2.2.1. <teiHeader>

The <teiHeader> should provide a relatively complete bibliographic record for both the electronic document and the original source manuscript. The data that describe the electronic text should be encoded in the <fileDesc> section, while data that pertain to the source text should be encoded within the <sourceDesc> section (a subsection of <fileDesc>). Projects should be careful to describe their source documents accurately; if encoders have access only to photocopies rather than original manuscripts, that fact should be noted in <sourceDesc>.

Much of the information in a TEI header is normally pulled from the source document. Manuscripts, however, often do not contain title pages or formal responsibility statements. Therefore, both <fileDesc> and <sourceDesc> may, if the project chooses, contain supplied (and/or regularized) information only. There is no need to use brackets to indicate what information has been supplied, except in the case of <title> in the <sourceDesc><titleStmnt>. Titles that have been supplied in <sourceDesc> should be enclosed in brackets. The <title> in <fileDesc><titleStmnt>, because it is inherently supplied and applies solely to the electronic file, need never be enclosed in brackets.

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 sample TEI header below offers a general template for manuscript headers. More detailed instructions on elements that require special attention follow.

```

<teiHeader type="CDL/TEI text type">
  <fileDesc>
    <titleStmt>
      <title>TITLE OF TEXT : ELECTRONIC VERSION</title>
      <author>LASTNAME, FIRSTNAME, YYYY-YYYY</author>
      <editor>LASTNAME, FIRSTNAME, YYYY-YYYY </editor>
      <respStmt>
        <resp>Text encoder:</resp>
        <name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name>
      </respStmt>
    </titleStmt>
    <editionStmt>ELECTRONIC EDITION STMT</editionStmt>
    <extent>in Kb. or Mb.</extent>
    <publicationStmt>
      <publisher>NAME OF PUBLISHER</publisher>
      <pubPlace>PLACE OF PUBLICATION</pubPlace>
      <date reg="YYYYMMDD">DATE OF PUBLICATION</date> [or date of creation of e-text? a
      <idno>OAC ARK: </idno>
      <idno>Local ID#</idno>
    </publicationStmt>
    <seriesStmt>
      <title>SERIES TITLE</title>
      <respStmt>
        <resp></resp>
        <name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name>
      </respStmt>
      <idno>OAC ARK: </idno>
    </seriesStmt>
    <sourceDesc>
      <bibl>
        <title>[TITLE]</title>
        <author><name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name></
        <respStmt>
          <resp>MS revisions by:</resp>
          <name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name>
          <resp>Amanuensis:</resp>
          <name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name>
        </respStmt>
        <extent># of MS leaves</extent>
        <date value="MMDDYYYY">DATE OF CREATION OF MS</date>
        <idno type="LOCAL">LOCALLY DEFINED ID#</idno>
      </bibl>
    </sourceDesc>
  </fileDesc>
  <encodingDesc>
    <projectDesc>
      <p> </p>
    </projectDesc>
    <editorialDecl>
      <p></p>
    </editorialDecl>
    <refsDecl>
      <p></p>
    </refsDecl>
    <classDecl>
      <taxonomy id="CONTROLLED VOCAB ID #1">
        <bibl>
          <title>NAME OF CONTROLLED VOCAB</title>
        </bibl>
      </taxonomy>
      <taxonomy id="CONTROLLED VOCAB ID #2">
        <bibl>

```

```

        <title>NAME OF ANOTHER CONTROLLED VOCAB</title>
    </bibl>
</taxonomy>
<taxonomy id="CONTROLLED VOCAB ID #3">
    <category>
        <catDesc>NAME OF LOCAL OR PROJECT DEFINED CONTROLLED VOCAB</catDesc>
    </category>
</taxonomy>
<classDecl>
</encodingDesc>
<profileDesc>
    <creation>
        <date></date>
    </creation>
    <langUsage>
        <language id="el">(Range: 0370-03FF)</language>
    </langUsage>
    <handList>
        <hand id="UNIQUE ID#" first="YES OR NO"/>
    </handList>
    <textClass>
        <keywords scheme="ID# FROM ONE OF THE <TAXONOMY ID="">'S">
            <list>
                <item>NAME OR SUBJECT HEADING</item>
                <item> </item>
            </list>
        </keywords>
        <keywords scheme="ID# FROM A SECOND <TAXONOMY ID="">">
            <list>
                <item>NAME OR SUBJECT HEADING</item>
                <item> </item>
            </list>
        </keywords>
    </textClass>
</profileDesc>
<revisionDesc>
    <change>
        <date>Date of revision</date>
        <respStmt>
            <name>PERSON RESPONSIBLE FOR REVISION</name>
        </respStmt>
        <item>EXPLANATION OF REVISION</item>
    </change>
</revisionDesc>
</teiHeader>

```

2.2.1.1. <fileDesc>

The <fileDesc> describes the electronic file in detail. For manuscripts, the following elements within <fileDesc> require special attention not otherwise required or emphasized by the CDL TEI Header guidelines:

2.2.1.1.1. <title>

The <title> in <fileDesc><titleStmt>, because it includes a subtitle ("electronic text") that is always supplied by the encoder and applies solely to the electronic file, need never be enclosed in brackets even if the main part of the title is supplied. Projects producing many files with supplied titles should spell out their procedures for supplying those titles in the <editorialDecl>.

For letters, the title should contain the names of the sender and the recipient and the date the letter was written.

```
<title>Letter from Samuel L. Clemens to Olivia L. Clemens,
29 July 1879: an electronic text</title>
```

Again, The title above title is not given in brackets because it refers to the electronic text. The same supplied title for the source document (<sourceDesc><title>) would be given in the following form.

```
<title>[Letter from Samuel L. Clemens to Olivia L. Clemens, 29 July 1879]</title>
```

[P4: 5.2.1]

2.2.1.1.2. <respStmt>

The <respStmt> must include the names of all document encoders, editors, and transcribers responsible for the electronic text. Individuals responsible for proofreading or making general corrections to the encoded file may also be recorded. Group here any other individuals or organizations that have not been entered using other elements (<author>, <editor>, etc.) but that have contributed to the work.

Each <respStmt> should contain at least one pair of <resp> and <name> elements. The <resp> element should contain a statement that indicates the nature of the contribution. Each <name> element should contain the name of the contributor. All documents should have at least one <respStmt> for the individual responsible for encoding the electronic text and the individual responsible for transcribing the manuscript. See the CDL TEI Header guidelines for instructions regarding the regularization of names. Within <respStmt>, repeat <resp> and <name> as necessary. Multiple names can be paired with a single <resp>, and multiple responsibilities can be described in a single <resp>.

```
<titleStmt>. . .
  <respStmt>
    <resp>Transcription by<resp>
    <name reg="Steadman, Ralph">Ralph Steadman</name>
    <resp>Encoded by</resp>
    <name reg="Rios, Leigh">Leigh Rios</name>
    <resp>Proofing and Corrections to markup by</resp>
    <name reg="Payne, Charlotte">Charlotte Payne</name>
  <respStmt>
</titleStmt>
<titleStmt>. . .
  <respStmt>
    <resp>Transcription by<resp>
    <name reg="Steadman, Ralph">Ralph Steadman</name>
    <resp>Encoded by</resp>
    <name reg="Rios, Leigh">Leigh Rios</name>
    <resp>Proofing and Corrections to markup by</resp>
    <name reg="Payne, Charlotte">Charlotte Payne</name>
  <respStmt>
</titleStmt>
```

[P4: 5.2.1]

2.2.1.2. <sourceDesc>

As unpublished documents, manuscripts can pose a problem for formalized bibliographic citation using <biblFull>. Therefore these guidelines recommend the use of <bibl>, which can accommodate totally unstructured bibliographic information. Within <bibl>, many of the elements available in <fileDesc> may be used.

```
<sourceDesc>
  <bibl>
    <title>[TITLE]</title>
    <author><name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name></
  <respStmt>
    <resp>MS revisions by:</resp>
    <name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name>
    <resp>Amanuensis:</resp>
    <name reg="LASTNAME, FIRSTNAME, YYYY-YYYY">FIRSTNAME LASTNAME</name>
    </respStmt>
    <extent># of MS leaves</extent>
  <date value="MMDDYYYY">DATE OF CREATION OF MS</date>
  <idno type="LOCAL">LOCALLY DEFINED ID#</idno>
  </bibl>
</sourceDesc>
```

[P4: 5.2.7]

2.2.1.2.1. <respStmt>

Transcribe responsibility statements if necessary, even if they are not explicitly given by the source document. For instance, a manuscript that has been extensively revised or annotated by someone other than the author may have the reviser's or annotator's name recorded. Or it may be important to record the name of an amanuensis in the <sourceDesc><bibl> even though that person's participation in the creation of the original manuscript may not warrant recording in the <fileDesc>. Statements that describe the nature of the intellectual or artistic contribution should be encoded in <resp>. Names of people or organizations should be placed in <name>. Repeat <resp> and <name> for each contributor. Please see the complete CDL TEI Header guidelines for when to use the **reg** attribute in <name> within <respStmt>.

```
<respStmt>
  <resp>Revised by</resp>
  <name reg="Paine, Albert Bigelow">Albert Bigelow Paine</name>
</respStmt>
```

2.2.1.2.2. <extent>

The <extent> element in <sourceDesc><bibl> will record the number of manuscript leaves or other relevant measure (e.g., pages) of the length of the original document.

2.2.1.2.3. <date>

The <date> in <sourceDesc><bibl> should normally reflect the date of primary creation of the manuscript, not the date of subsequent revision. Where significant revision has occurred over time and must be recorded, date ranges may be given within <dateRange> instead.

2.2.1.3. <encodingDesc>

Through a variety of sub-elements, the mandatory <encodingDesc> gives information about the editorial and encoding policies behind the electronic text.

2.2.1.3.1. <editorialDecl>

All manuscript projects must also use the <editorialDecl> element to describe their transcription policies, especially those regarding non-typographical elements, editorial commentary, and revisions. This is where projects must also describe all policies regarding the emendation of the source text, including correction of apparent errors, standardization of spelling, expansion of abbreviations, exclusion of certain elements, etc.

[P4: 5.3.3]

2.2.1.4. <profileDesc>

2.2.1.4.1. <handList>

Projects working with manuscript that is written by more than one person may wish to distinguish between the hands that created the document. Use <handList> to identify all hands that appear in the document, listing each hand using the <hand> element. Usually, the names associated with those hands will also then be recorded in the <respStmnt> of the <fileDesc> and <sourceDesc>.

2.2.1.4.2. <hand>

The <hand> element is an empty tag whose attributes are used to describe each distinct hand in the document. The following attributes are required:

value	description
id	gives a unique identifier for the hand
first	"yes" if hand is primary or dominant hand in the document, "no" if it is not

The **scribe** attribute may also be used to give another name or non-unique identifier for the hand.

```
<handList>
<hand id="SLC" scribe="Clemens, Samuel Langhorne" first="yes"/>
<hand id="ABP" scribe="Paine, Albert Bigelow" first="no"/>
</handList>
```

[P4: 18.2.1]

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

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 text being encoded. The <text> element contains three sub-elements, <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>]

2.3.2. <group>

Manuscript documents can present a special challenge for grouping. As far as it is possible, documents should retain structural integrity. Those that cannot stand alone content-wise should not stand alone as files. Therefore, while individual letters in a series of correspondence may be offered as separate TEI documents, the manuscript of a novel should normally be contained within one file, with appropriate divisions for chapters, etc. Working notes for that novel, however, could be encoded separately. Individual texts that may stand alone but whose grouping is deemed significant enough to be worth retaining may also be encoded as a group. Therefore, a collection of poetry or short stories that has been grouped and ordered by the author may be encoded in a single file, or, if the author has not ordered them in a discrete group, may be separately encoded.

When grouping documents, bear in mind, too, that all documents that are grouped in a single file will have the same TEI header. If you wish to have very granular control of your documents, you may wish to give each one its own header, and therefore encode each one as a separate file.

Normally, the <divn> element will be enough to create the structural divisions necessary for documents that are encoded together as a group. However, for groups of texts for which the <divn> structure is not enough, for instance for a collection of short stories in which each manuscript text has its own distinct front matter, the <group> element may be used to contain multiple <text>s. Avoid using <teiCorpus>.

```
<TEI.2>
  <teiHeader> . . . </teiHeader>
  <text>
    <front>
      <titlePage> . . . </titlePage>
    </front>
    <group>
      <text>
        <front>
          <titlePage> . . . </titlePage>
        </front>
        <body> . . . </body>
        <back> . . . </back>
      </text>
      <text> . . . </text>
      <text> . . . </text>
```

```

        <text> . . . </text>
      </group>
    </text>
  </TEI.2>

```

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

2.4. Front Matter

2.4.1. <front>

The <front> element is used to contain the various components that make up front matter, including prefaces, introductions, dedications, and title pages. Each of these sections is normally contained within another structural element such as <titlePage> or a <divn>. For a full list of the types of <divn>s available in <front>, please see the section on divisions. Note that <front> should be reserved for containing prefatory material that is relevant to the entire body text *as a whole*. Other section titles or attributive information that happens to occur at the beginning of the document but refers to only a portion of the document (e.g., the title of the first poem in a short collection of poems) should be contained within appropriate elements within the appropriate <divn>.

2.4.2. <titlePage>

Do not encode a <titlePage> unless the manuscript itself contains a formal title page. As with any page in a manuscript, title pages may often possess a number of formatting peculiarities, such as specific alignment, fonts, incidental images, etc. It is not necessary to attempt to reproduce the look or layout of the title page exactly. It is often enough to convey to users what textual information the title page contains and the order in which the information appears.

```

<titlePage>
  <docTitle>
    <titlePart type="main">The Mad Mariner</titlePart>
    <titlePart type="subtitle">Being a Tale of Harrowing Heroism</titlePart>
  </docTitle>
  <docAuthor><name>Wally Bunham</name></docAuthor>
  </docImprint>
</titlePage>

```

2.4.2.1. <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">The Long Way Home</titlePart><titlePart type="subtitle">My
</docTitle>

```

2.4.2.2. <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>
```

2.4.2.3. <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>
```

2.4.2.4. <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>
```

2.4.2.5. <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. For a full list of the types available, please see the section on divisions.

[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 afterwords, epilogues, 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. For a full list of the types of <divn>s available in <back>, please see the section on divisions.

```
<back>
  <div1 type="epilogue">
    <head>Epilogue</head>
    <p>And so it came to be that all the creatures. . .
    </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">
```

The following **type** attribute values are suggested for manuscripts. Please feel free to create other **type** values as needed, remembering to document them in the <editorialDecl>. Please note that the **types** listed below may be used for <divn>s in <front>, <body>, or <back> as necessary.

For correspondence:

value	type of division
letter	main letter
letText	letter text (as opposed to letterhead, etc.)
enclosure	enclosure
envelope	envelope
letPart	letter part
postscript	postscript
letterhead	letterhead
secLet	secondary letter
docket	docket
postmark	postmark
addressee	name and/or address of recipient
sender	name and/or address of sender

For literary manuscripts

value	type of division
chapter	chapter
section	section
ss1-ss6	sections 1-6
verse	poem or other verse (see also 2.11 on verse types)

value	type of division
introduction	introduction
epilogue	epiglogue
comment	commentary
part	manuscript part
dedication	dedication

Manuscripts that are formatted like books may also borrow from the div types available in the CDL TEI Printed Book guidelines.

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

2.8. Division Headings, Openers, and Closers

Significant textual divisions often open with a heading identifying the content of the division. They may also begin and end with phrases such as bylines, epigraphs, datelines, and the like.

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>
```

2.8.2. <epigraph>

Epigraphs contain quotations, anonymous or attributed, appearing at the start of a section, chapter, or other major division. They should be enclosed within the <epigraph> element. An epigraph appearing on a page by itself should be encoded in <epigraph> within a <divn type="epigraph">.

Within <epigraph>, attributed epigraphs should be enclosed entirely within the <cit> element, with <quote> containing the quoted passage and <bibl> containing the attribution. Within <quote>, use <p>, <lg>, or other block elements as necessary.

```
<epigraph>
  <cit>
    <quote>"I believe that any other ideal is impracticable and is a collision with huma
    and God."</quote>
    <bibl>Attributed to George Herron.</bibl>
  </cit>
</epigraph>
```

```

<epigraph>
  <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>
</epigraph>

```

Within `<epigraph>`, unattributed epigraphs should simply be encoded within `<quote>`, with `<p>` and other block elements used as necessary to contain the quoted passage. There is no need to use `<cit>` for unattributed epigraphs.

```

<div1 id="ch01" type="chapter" n="1">
  <head>Chapter 1</head>
  <epigraph>
    <quote>
      <p>I pity the man who can travel from Dan to Beersheba<p>
    </quote>
  </epigraph>

<epigraph>
  <quote rend="italic">
    <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>
</epigraph>

```

2.8.3. `<opener>`, `<closer>`

Projects encoding correspondence or other letter-like documents may wish to use `<opener>` and `<closer>` (and other elements that are allowed within them such as `<salute>`, `<signed>`, and `<dateline>`) in order to take advantage of the additional structuring they provide. They allow for certain text features that need to be identified for analytical or formatting purposes--such as, addresses, datelines, salutations, signatures, or postscripts--to be grouped neatly at the beginning or end of `<divn>`s. Because these elements can require fairly heavy analytic encoding and introduce additional encoding restrictions, they are not required. In fact, they are discouraged for projects that do not need to perform a high level of analysis with their encoding. Projects wishing to use `<opener>` and `<closer>` must consult TEI P4 for more detailed instructions on their use. The letter containing the following text could be encoded using different structural hierarchies, depending on what text features the project wishes to make distinct:

From the Desk of Martin Snope

The Willows

Chicago, Illinois

June 14, 1952

Dear Isabella,

I have arrived home at last and will not leave again until August.

Encoded with <opener>:

```

    <text>
  <body>
  <div1 type="letter">
  <opener>
  From the Desk of Martin Snope
  <address><addrLine>The Willows</addrLine>
  <addrLine>Chicago, Illinois</addrLine>
  <address>
  <dateline>June 14, 1952</dateline>
  <salute>Dear Isabella,</salute>
  </opener>
  <p>I have arrived home at last . . . </p></div1>

```

Encoded without <opener>:

```

  <text>
  <body>
  <div1 type="letter">
  <p>
  From the Desk of Martin Snope
  <address><addrLine>The Willows</addrLine>
  <addrLine>Chicago, Illinois</addrLine>
  <address>
  </p>
  <p>June 14, 1952</p>
  <p>Dear Isabella,</p>
  <p>I have arrived home at last . . . </p></div1>

```

2.8.4. <byline>

Bylines are formal statements of responsibility, which may sometimes be found near the top of a division (usually after a <head>) and sometimes at the bottom. Do not use <bylines> to record attributive information for quoted passages; use instead the <cit>/<quote>/<bibl> structure described in the section on quotations. Do not use <byline> for the attribution of correspondence, which is normally signed (<signed>). Do not use <byline> when a more complete bibliographical citation is present; in that case <bibl> is normally more appropriate. (See the section on bibliographic citation.) Take care not to confuse the the use of <byline> and similar elements within <divn>s with their use within formal <titlePage>s.

```

<div1 type="introduction">
  <head>Introduction</head>
  <byline>by Sherna Gluck</byline>
  <p>The following interviews with Sylvie Thygeson represent two distinct interviews ...

```

```
<div2 type="essay">
  <head>In the Public Interest—Jeannette Rankin</head>
  <bibl>by <author>Ralph Nadar</author>
    (<title rend="italic">The New Republic Feature Syndicate</title>
     <biblScope>Number 33</biblScope>
     <date>September 11, 1972</date>)
  </bibl>
  <p>WASHINGTON—A few weeks ago we sent a questionnaire ...
```

2.8.5. <dateline>

Use <dateline> to encode a place and date associated with the creation of the document. Encode the place name directly within <dateline>, but use <date> to enclose the date itself within <dateline>. When additional address information is available, use <address> within <dateline>. (See the section on addresses.) As with <byline>, do not use <dateline> to encode more complete bibliographic citations. Use <bibl> instead.

Example:

```
<div1 type="chapter">
  <head>Prologue</head>
  <dateline>March 1945: Shensi Province, China</dateline>
  <p>A dull orange haze, the first light of dawn, ...
```

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 manuscript (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/> between <divn> elements. All such interstitial page numbers should be encoded as if they belonged to the nearest subsequent <divn>, before any <head> elements. If a page break occurs in the middle of a smaller block element (e.g., <p>), it can simply be encoded there.

The **n** attribute should be used to record the physical position of the page within the sequence of pages. Sometimes this numbering will differ from the numbers that the manuscript author has provided. (Page numbers written on the manuscript are recorded within <fw>. See below.)

Projects may choose to number manuscript pages or leaves in a variety of ways. If the text of a document primarily appears on one side of a leaf, then the recto pages may be numbered 1, 2, 3, etc. with the occasional text on the verso labeled 1verso, 2verso, 3verso, etc. If the text is written on both sides of the leaf, then each side may be considered its own page and numbered 1, 2, 3, 4, etc. If the document has been folded so that text runs in columns on the physical leaf and it is necessary to differentiate the two columns as separate "pages" (because the author did so or because each column was scanned into a separate image file), then each column may also be numbered 1, 2, 3 and separated with page breaks. Regardless of which numbering scheme a project chooses, empty pages need not be labeled as such; a <pb/> followed directly by another <pb/> will suffice.

```
<pb n="1" />
<pb n="1verso" />
<pb n="2" />
```

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

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" />
```

2.10.1.1. <fw>

Catchwords and page numbers may be recorded in <fw> ('forme work'). These need not necessarily be recorded unless full diplomatic transcription standards are being pursued or projects otherwise find it important to record them. Nevertheless, it is generally good practice to record in <fw type="pag"> the page number assigned by the author in order to preserve the original order and/or understand where there may be gaps or additions to the manuscript. Use <fw> only for numbers that are actually written on the page:

```
<fw type="pag">1</fw>
```

The `<fw>` element must contain a **type** attribute. Possible values are:

value	description
header	a running title at the top of the page
footer	a running title at the bottom of the page
pag	a page number or foliation symbol
sig	a signature or gathering symbol
catch	a catch-word

Page numbers in manuscripts have often been revised. Projects wishing to capture such revision may do so within `<fw>` using `<add>` and ``. See the section on revisions for more information.

```
<fw type="pag"><del>1</del><add>2</add></fw>
```

The `<fw>` element should directly follow the `<pb>` element indicating the start of the page on which it appears, regardless of where it actually physically appears on the page. Projects that wish to record the location of the content of `<fw>` should use the **place** attribute.

[P4:18.3]

2.10.2. `<lb>`

Normally, full diplomatic transcription would require that the text be transcribed line-for-line and the line breaks recorded in XML using the empty `<lb/>` tag to mark the start of a new line. However, projects may forgo this practice, as long as the policy regarding transcription of lineation is expressly set out in the `<editorialDecl>`.

Projects should be aware that not transcribing manuscripts line-for-line may create difficulties in transcription that may only be resolved by emendation (e.g., hyphenation). Therefore, they should only forgo recording line breaks when they feel prepared to develop and consistently apply a specific emendation policy, possibly also involving the use of TEI's critical apparatus mechanism. Any such policy must be documented in the `<editorialDecl>`.

Note that the `<lb>` tag is intended for marking line breaks in prose only. The `<l>` element must be used to mark lines of verse. (See the section on verse.)

```
<p>The next Saturday, we went<lb/>
to the park and sat feeding<lb/>
the pigeons.</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. 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#COR55>]

2.11. Typographical Phenomena and Formatting

2.11.1. <hi>

Record underscoring and other typographical highlighting with the <hi> element. Use the required **rend** attribute to record the type of highlighting 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.

When text with special highlighting has already been tagged for other structure or content, and when the special highlighting 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 small caps, 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 document; unemphasized or highlighted 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
overline	written with a line drawn above the text
strikethrough	strikethrough
subscript	below the baseline of standard text

value	display
superscript	above the baseline of standard text
hide	do not display

Additional suggested rend values for manuscripts are:

value	display
fancy	cursive or stylized text
preprint	pre-printed or typeset text (e.g., on stationery)
2-underline	double underscore
3-underline	triple underscore
4-underline	quadruple underscore
del-underline	deleted underscore
2-del-underline	deleted double underscore
3-del-underline	deleted triple underscore
4-del-underline	deleted quadruple underscore
wavy-underline	wavy underline
dash-underline	dashed underline
dot-underline	dotted underline
paraph	paraph or flourish beneath text
boxed	text that has a box drawn around it
circled	text that has a circle drawn around it

Text that is not different in style but in orientation may also be distinguished using the **rend** attribute. Available **rend** values are:

value	display
cross-written	written across previously written text
diag	written diagonally
vert	written vertically, at 90 degrees to rest of text
up-down	written upside-down on the page

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
```

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 typographic highlighting that explicitly conveys emphasis rather than some 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, standard initial indent
center	center
right	justify right, ragged left, standard initial indent
justify	fully justify
indent	standard paragraph indent
hang	hanging indent
blockindent	full block indent

value	display
blockquote	full block indent used for quotes (<quote>)
noindent	fully flush left, with no initial indent
inline	keeps elements that usually create a line break from starting a new line; only the element that would follow the break needs to be tagged as 'inline'
halfline	records a half line space before the tagged element

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. Changes in Hand

Projects may wish to record in their encoding the fact that some parts of the text are written in a hand other than that of the primary author. To do so, projects must first include in the <handList> of the TEI header all <hand>s that are present in the manuscript. (See the section on <handList> above.) Recall that each <hand> must carry a **first** attribute that will indicate whether that hand is considered the primary hand in the manuscript or not. All elements will be considered as having been written in the primary hand unless otherwise encoded.

To encode the presence of another hand, use the **hand** attribute in an element that encompasses the text in question. The value of the **hand** attribute must correspond to the **id** attribute of the appropriate `<hand>` in the `<handList>`. The **hand** attribute is available in the following elements:

```
<p>
<l>
<note>
<add>
<add>
<del>
<ab>
<seg>
```

The `<ab>` and `<seg>` elements are arbitrary containers that can be used when there is no other appropriate encompassing tag on which to hang a **hand** attribute.

```
<p>Where <del hand="AC">the heck</del> did they go?</p>
```

2.14. 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>
```

2.14.1. `<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>

```

2.14.2. <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>]

2.15. Speech

Texts that are made up primarily of attributed speech--e.g., plays, screenplays, and interview transcripts--should be encoded using the <sp> and <speaker> elements. Transcriptions of speech embedded in prose or verse texts may also be encoded using these elements.

2.15.1. <sp>

The <sp> element is used to contain instances of speech in a performance text or a transcript of spoken words in a prose or verse text. The entire speech along with its attribution should be encoded within <sp>. Within <sp>, use <p>, <lg> and other block elements as necessary to format and contain the contents of the speech.

```
<sp>
```

```

<speaker>FILCH.</speaker>
<p>Sir, Black Moll hath sent word her Trial comes on in
  the Afternoon, and she hopes you will order Matters
  so as to bring her off.</p>
</sp>
<sp>
  <speaker>PEACHUM.</speaker>
  <p>Why, she may plead her Belly at worst; to my
    Knowledge she hath taken care of that Security.
    But, as the Wench is very active and industrious,
      you may satisfy her that I'll soften the Evidence.</p>

```

The `<sp>` element may also carry an optional **who** attribute that gives the identity of the speaker. The value of **who** must refer to the **id** of a person previously identified in either a cast list (`<role>` in `<castItem>` in `<castList>` for dramas and screenplays) or a description of the participants in a transcribed speech or interview (`<person>` in `<partiDesc>`). For more specific information on how to assign **ids** that would be valid in **who**, see P4: 10.1.4 [<http://www.tei-c.org/P4X/DR.html#DRCAST>] for encoding cast lists, P4: 5.4 [<http://www.tei-c.org/P4X/HD.html#HD4>] for encoding participants.

```

<profileDesc>
  <particDesc>
    <person id="LaBerge" role="interviewer/editor">
      <persName reg="LaBerge, Germaine">Germaine LaBerge</persName>
    </person>
    <person id="Bouche" role="interviewee">
      <persName reg="Bouché, Briec">Briec Bouché</persName>
    </person>
  </particDesc>
  ...
  <sp who="LaBerge">
    <speaker>LaBerge</speaker>
    <p>Why don't we start with where you were born, and a little bit about your family b
  </sp>
  <sp who="Bouche">
    <speaker>Bouché</speaker>
    <p>Yes. How much detail do you want? Full detail or just very sketchy?</p>
  </sp>

```

2.15.2. `<speaker>`

The `<speaker>` element is used within `<sp>` as a specialized form of heading giving the name of the speaker responsible for the spoken words. Encode the name of the speaker as it is given in the source document. Do not supply a name if one does not appear in the source text. The content of `<speaker>` is displayed in bold and flush left on a line preceding the text of the speech.

```

<sp who="LaBerge">
  <speaker>LaBerge</speaker>
  <p>Why don't we start with where you were born, and a little bit about your family b
</sp>
<sp who="Bouche">
  <speaker>Bouché</speaker>
  <p>Yes. How much detail do you want? Full detail or just very sketchy?</p>

```

```
</sp>
```

is displayed as:

LaBerge

Why don't we start with where you were born, and a little bit about your family background?

Bouché

Yes. How much detail do you want? Full detail or just very sketchy?

Projects that require a different kind of styling for the display of speaker names should use the **rend** attribute to override the default styling imposed by `<speaker>`.

[P4: 10.2.2] [<http://www.tei-c.org/P4X/DR.html#DRSP>]

2.16. Verse

2.16.1. `<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.)

2.16.2. `<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">
```

2.16.3. `<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">
```

2.16.4. `<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>

```

2.16.5. `<closer>`

Occasionally a poem will be directly followed by a date or some other closing information that is not considered part of the poem. This information may be encoded using the `<closer>` element within the poem's `<divn>` outside of the last `<lg>`. Poems must be enclosed within a `<divn>` in order to use `<closer>`.

```

<div1 type="poem">
. . .
<lg type="stanza">
<l>Nor certitude, nor peace, nor help for pain;</l>
<l>And we are here as on a darkling plain</l>
<l>Swept with confused alarms of struggle and flight,</l>
<l>Where ignorant armies clash by night.</l>
</lg>
<closer>[1867]</closer>
</div1>

```

[P4: 6.11] [<http://www.tei-c.org/P4X/CO.html#CODV>]

2.17. Notes

2.17.1. `<note>`

Use the <note> element to encode notes, using the **place** attribute to indicate the location of the note. According to TEI P4, a "note is any additional comment found in a text, marked in some way as being out of the main textual stream." There are three kinds of notes in manuscript: authorial notes, commentary from another person writing on the original document, and editorial notes from the transcriber/encoder producing the electronic text. For convenience, we will call only the last kind "editorial." This section deals exclusively with authorial notes and other commentary that is *not* editorial. (Editorial notes are discussed in the section on editorial intervention.) Authorial notes and non-authorial commentary are generally distinguished as notes typographically on the page, either through different hand/ink or by placement in the margins or by being written in a different orientation. Text that is not distinguished typographically need not necessarily be treated as notes.

2.17.1.1. Note Place

Notes should be tagged in-line or in proximity to the text to which it is relevant. A marginal note referring to something in the manuscript, should be tagged directly adjacent to the text that it refers to, either preceding or following it depending on the order in which the text and note are meant to be read. The **place** attribute should be used to describe the physical location of the note on the page. The value given in the **place** attribute will also be used to dictate the display of notes where possible. Any note without a place value will by default be treated as in-line and displayed embedded where it has been tagged in the text. Sample values for the **place** attribute are:

value	note place
foot	note appears at foot of page.
end	note appears at end of chapter or volume.
inline	note appears as a marked section in the body of the text.
left	note appears in left margin.
right	note appears in right margin.
top	note appears in top margin
interlinear	note appears between lines of the text.
verso	note appears on back side of page on which related text appears
recto	note appears on front side of page on which related text appears

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 content of notes in <p>s or other appropriate block elements if necessary. (See the section on internal linking for more information about <ref>s.)

2.17.1.2. Note Attribution

The **resp** attribute is mandatory for all notes to indicate who is responsible for the annotation: author, editor, translator, etc. Available values for the **resp** attribute are:

value	note attribution
auth	note originated with the author of the text
rev	note added by a reviser of the text who has written on the document
ed	note added by the editor of the text
comp	note added by the compiler of a collection

value	note attribution
tr	note added by the translator of a text
transcr	note added by the transcriber of a text into electronic form. (See 2.15.1.)
amanuensis	note added by the amanuensis
(other id)	note added by the individual indicated by the initials or other unique id; id must be defined as the id of a <hand> in the TEI header

2.17.1.3. Note References

Notes that depend on a reference system can be automatically linked to references using the **n** attribute in <note>. Numbers, asterisks, dingbats, crosses, daggers, and the like, are recorded as the value of the **n** attribute and need not be included in the text as a <ref>. UNICODE characters may also be recorded in **n** attribute when they serve as references. A <note> does not require a unique **id** attribute unless it is targeted by a <ref>.

```
<p>...when I returned to the garden, she had already gone.</p>
<note resp="auth" place="left" n="*">I remember now that she must have headed for the sh
<p>Not knowing what to do, I ...</p>
```

2.17.2. 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>
```

2.17.3. Footnotes and Marginal Notes

Footnotes (those references, notes, and citations appearing at the bottom of the page) and marginal notes must be encoded where they are referenced. In other words, at the location of the note reference in the text, embed the <note> itself in place. If a footnote or marginal note 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 at 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 note from other locations in the text may be tagged with <ref>. If a note is targeted by any <ref> anywhere in the text, it must include an **id** attribute.

```
<p>There is no evidence whatsoever that anyone could possibly have know about it.
<note id="note1" place="left" n="*" hand="AB">
<p>There is in fact ample evidence now in the pubic record.</p></note>...</p>
```

2.17.4. 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 <div> 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>]

2.18. Editorial Intervention

2.18.1. Editorial Notes

Editorial notes are defined here as those that are provided by the transcriber/encoder of the document and that do not appear in the source document itself. Editorial notes should be tagged in-line as <note>s, adjacent to the text to which they are relevant. They should be reserved for instances in which no other method of tagging is adequate to describe or tag the textual situation. While they can be used to qualify or describe text that is in the original, they should not be used to supply conjectural text or corrections or variant readings. TEI P4 supplies other mechanisms for those type of editorial interventions. Please consult TEI P4 for further discussion.

All editorial notes must have a **resp** attribute identifying the agent responsible for the note, in part to distinguish them from authorial notes. Available resp values are the same as those for authorial notes, as long as it is absolutely clear whether the note is editorial or authorial. If for instance, there are two compilers--one who has written on the document itself and the other who has grouped the documents only for electronic publication and has offered some necessary commentary--projects will wish to come up with different resp values for each person:

```
<note resp="ms-comp"> or <note resp="ed-comp">
```

In display, editorial notes will be distinguished from the rest of the text through brackets or some other exclusive editorial sign. The brackets or typographical signs are automatically created by a stylesheet when it recognizes a **resp** value as being editorial. There is therefore no need to put brackets around the content of the <note> during tagging.

[P4: 6.8.1]

2.18.2. <sic>, <orig>

In their transcriptions, some projects may wish to correct apparent errors in the text or to regularize non-standard spelling or punctuation. Any policy calling for such changes to the source text must be explicitly spelled out in the <editorialDecl>. If such a policy is consistently applied and transparent, then it may not be necessary to mark each and every location in the text where changes have been made. If all ampersands, for instance, are to be changed to "and", then that global change can be declared in <editorialDecl> and the changes themselves made silently.

However, if projects feel that it is necessary to correct or standardize text in individual instances and they would also like to preserve the text of the original, they may use the <sic> and <orig> elements.

Use the <sic> element in conjunction with its **corr** attribute to provide the text as it appears in the manuscript and the correction that needs to be provided.

```
<sic corr="tomorrow">tomorrow</sic>
```

The <sic> element may also be used without a **corr** attribute simply to call attention to apparent errors in the text.

The <orig> element works in a similar manner. Use <orig> to standardize non-standard spelling, punctuation, or syntax. Surround the original text with <orig> and use its **reg** attribute to contain the regularized version of the text.

```
<orig reg="don't">dont</orig>
```

[P4: 6.5.1]

2.19. Gaps, Illegible Text, Damage

2.19.1. <gap>, <unclear>

TEI P4 provides much discussion of how to transcribe problematic textual situations such as gaps and damaged manuscript. All projects transcribing manuscripts are encouraged to consult Chapter 18 of TEI P4 for a more thorough discussion of these features. The section 18.2.4 in particular contains a useful explanation of when and how to use the various elements and attributes available to record such phenomena.

Not all projects will need all the elements provided. Basic transcription projects should simply use the following elements for text that cannot be read because of damage or other obscuration:

Use the empty <gap> element at the point where the text is missing or completely illegible and no text is supplied to replace it.

Use the <unclear> element to enclose text that can be read but, because of damage or other reason, the reading is uncertain.

For both elements, the **reason** attribute may be used to state the cause (damage, deletion, etc.) of the loss of text. The **extent** attribute can also be used to indicate the extent of missing text.

Projects wishing to provide additional physical description of a document, may also use the <damage> element to enclose text that is fully legible but otherwise damaged in some way. Passages that are marked with <damage> can also contain <gap>s and <unclear> portions.

<p>The reason for his departure was <gap reason="torn"/> was not prone <damage reason="w

[P4: 6.5.3]

2.20. Revisions in Manuscripts

2.20.1. <add>,

Use <add> to encode additions/insertions in the text. An insertion is defined as text that has been placed between two previously inscribed words or characters, or between such a word or character and a previously fixed point (such as the top of the page), thus written later than the text on either side of it.

Use similarly to contain deleted content. If text has been added and then portions of it later deleted, the portion should be tagged as such within the text contained by <add>.

Note that content marked up using <add> and without additional linking attributes will be associated by proximity but will not be explicitly linked. Normally, tag the deletions and additions in the order that they appear. Projects interested in documenting explicit substitution (e.g., the content of <add> is a substitute for the content of) should consult the full TEI guidelines.

<p>Paul handed the cash<add>money</add> over to the old teller.</p>

[P4: 6.5.3]

2.20.2. Revisions in Another Hand

When <add>s or s have been made in a hand other than the main hand of the document, use the **hand** attribute to identify the writer. The value of the **hand** attribute must be identified as a <hand> in <handList> in the TEI header.

<p>The witness for the prosecution refused <add hand="AC">repeatedly</add> to answer the

2.21. Names, Dates, and Addresses

Although it is not required, it is sometimes useful to tag names, dates, and addresses as they occur throughout the text, not only when they occur on the title page. Tagging names and dates also allows them to be regularized in order to provide more fruitful searching.

2.21.1. <name>

The <name> element may be used to encode any proper noun or proper noun phrase. The **type** attribute can be used to

indicate the type of name. Supported **type** values are "person" and "place". The **reg** attribute may be used to give a normalized or regularized form of the name.

```
At the time of the events which led to
<name reg="Benedict XII, Pope of Avignon (Jacques Fournier)"
type="person">Fournier's</name> investigations,
the local population consisted of between 200 and
250 inhabitants.
```

2.21.2. <date>

Use <date> to encode a date that has been given in any format. The **value** attribute can be used to contain the value of the date in the standard ISO 8601:2000 5.2.1 format (e.g., YYYY-MM-DD). Again, this is useful if document dates need to be indexed for searching.

Because the <date> element is not directly allowed within <divn> it can be surrounded by <dateline> if necessary. When it appears at the beginning or end of a division, <date> is normally located within the <opener> or <closer> elements. Projects not wishing to use <opener> and <closer> may also insert <date> directly within <p> if that is appropriate.

```
<p>Given on the <date value="1977-06-12">Twelfth Day of June
in the Year of Our Lord One Thousand Nine Hundred and
Seventy-seven of the Republic the Two Hundredth and first
and of the University the Eighty-Sixth.</date></p>
```

2.21.3. <address>, <addrLine>

The <address> and <addrLine> elements can be used to encode postal or other addresses. Enclose the entire address within <address> and each individual line within <addrLine>.

```
<address>
  <addrLine>110 Southmoor Road,</addrLine>
  <addrLine>Oxford OX2 6RB,</addrLine>
  <addrLine>UK</addrLine>
</address>
```

Because <address> is not allowed directly in <divn>, when it appears at the beginning or end of a division, it normally is enclosed within the <opener> or <closer> elements. Projects not wishing to use <opener> or <closer> may insert <address> directly inside a <p> if that is appropriate.

```
<div1 type="letter">
  <head>Appendix: Letter to Earl Warren</head>
  <opener>
    <date>November 10, 1971</date>
    <address>
```

```

<addrLine>Honorable Earl Warren</addrLine>
<addrLine>Supreme Court of the United States</addrLine>
<addrLine>Washington, D. C.</addrLine>
<addrLine>Re: ACLU Proposed Earl Warren Civil Liberties Award</addrLine>
</address>
<salute>Dear Governor:</salute>
</opener>

```

[P4: 6.4] [<http://www.tei-c.org/P4X/CO.html#CONA>]

2.22. Lists

2.22.1. <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.

2.22.2. 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>
```

2.22.3. 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>]

2.22.4. `<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>
```

Occasionally, a `<list type="ordered">` in manuscripts will have enumerators that have been revised. In order to record those revisions, it is necessary to encode the sequential marker as the content of an element rather than as an attribute value. The revised enumerator is then encoded in `<label>`, directly preceding each `<item>`. To prevent duplication of the marker, encode the list with a `rend` attribute value of "label".

```
<list type="ordered" rend="label">
  <label>1.</label><item>Apples</item>
  <label><del>1.</del>2.</label><item>Oranges</item>
  <label><del>2.</del>3.</label><item>Grapes</item>
</list>
```

2.23. Bibliographies

2.23.1. `<bibl>`

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 annotations. 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>
```

2.23.2. <title> levels

Projects using the <title> element may also use its **level** attribute to define the type of title being provided and dictate the 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 item published as part of a larger item)	surrounded in quotation marks
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

2.23.3. <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>]

2.24. Internal Links and Cross References

Internal references and links can take many forms: numbers in the text that point to notes, pointers to specific sections of the text (e.g., "see the next chapter"), or short form bibliographic references (e.g., "Clemens 1890"). 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.

2.24.1. <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:

value	type of reference
citeref	bibliographic citation reference
figref	figure reference
fnoteref	footnote reference
formularef	formula reference
noteref	endnote or general 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 notes tagged at the point of reference is normally optional as the in-line presence of the <note> will automatically create a reference. References to the note from other locations are to be treated as <ref>s. See the section on footnotes and marginal notes 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>]

2.25. External Objects

2.25.1. <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 url pdf sound video stream	required
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>]

2.26. Graphic Elements

When encoding graphic elements such as illustrations, formulas, and tables, take special care to preserve both the information represented and, as far as possible, the form of presentation.

2.26.1. Tables

The CDL TEI Manuscript guidelines employ the full XHTML table module [http://www.w3.org/TR/2001/REC-xhtml-modularization-20010410/abstract_modules.html#s_tablemodule] instead of the TEI default table scheme to encode tables. See the full XHTML table module [http://www.w3.org/TR/2001/REC-xhtml-modularization-20010410/abstract_modules.html#s_tablemodule] guidelines for detailed instructions on how to encode tables. Projects should try as much as possible to encode for correct display in both Netscape and Internet Explorer browsers on the Windows and Mac platforms. (Take care to encode definition lists as `<list type="gloss">` when encountered; these can sometimes 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>
```

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

2.26.2. <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 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>]

2.26.3. Formulas

2.26.3.1. 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">
```

2.26.3.2. <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>]

2.27. 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.

2.27.1. <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"></addrLine>
  <addrLine>1515 42nd Street</addrLine>
  <addrLine>Chicago, IL</addrLine>
</address>
```

2.27.2. <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 <divm> 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>