

Guidelines for Markup of Electronic Texts

- **Last Updated:**2005-11-16
- **Editor:** Peter C. Gorman
- **Version:**2.0
- **Current Version:**
<http://uwdcc.library.wisc.edu/resources/etext/TEIGuidelines.shtml>

*Version 1.0 endorsed by the UW-Madison Libraries Digital Steering Committee
September 11, 2000*

Additions or changes made since this version was first adopted look like this.

1. Introduction

This document is intended for use by staff using the Text Encoding Initiative (TEI) Guidelines to mark up electronic texts for inclusion in the UW-Madison Libraries' digital collections. It is not relevant to other types of collections using SGML encoding, *e.g.*, e-facsimile collections or digital finding aids. Some of the content has been quoted or adapted from other published guidelines, which are referenced in each case.

The purpose of this document is not to teach or otherwise document the TEI itself, but rather to create a profile of the TEI for use in the UW-Madison digital library collections. *It is assumed that the user is already familiar with TEI markup.* The motivation for creating these guidelines is a desire to create a consistent and scalable infrastructure for text encoding projects, whereby new works can be created and added to the collection with minimal development effort on the part of project leaders, text encoders, and technical staff. At the same time, text encoded according to these guidelines should provide a suitable base for further elaboration or expansion by future encoders with minimal restructuring.

At any point in this document, you can click on the magnifying glass icon  to see examples of the point being discussed. The examples will open in a new window. If the meaning or use of any element is unclear, consult the TEI Guidelines for a full definition.

Terminology

These guidelines are *prescriptive*. Recommendations are made at three levels of emphasis:

- *must, must not, will, will not:* Unless the recommendation is followed, the document will not be considered valid at the level of encoding being described. Where possible, these recommendations will be enforced in a DTD.
- *should, should not:* The recommendation should be followed if at all possible; it should only be violated if the encoder has a good reason for doing so.

- *may*: The recommendation is suggested, but optional. Encoders may choose other (valid) strategies if they seem appropriate.

2. General Recommendations

- Texts following these Guidelines must be encoded as SGML (not XML) documents following the TEI P4 Guidelines and DTDs. 
- "Numbered <div>s present advantages to search and indexing software by explicitly communicating the hierarchical level of the section described. One anomaly of the TEI Guidelines is that <div0> is not available in <front> or <back> matter. Therefore, we recommend the use of numbered <div>s throughout the electronic text, always beginning with <div1>. Texts at all levels should include at least one <div1>." 

<div0> should be used, however, when the body of a work is organized at a level higher than that of chapter or article. For example, for a book consisting of "Part 1" and "Part 2", each of which contains chapters, the Parts should be encoded as <div0> and the Chapters as <div1>. 

- Because retrieval and navigation software commonly work with sequences of <divn> elements, all elements (including page breaks, <pb>) should occur within some <divn> unless logic or syntax prevents doing so. Otherwise, page breaks or other elements may be missing from user displays. <pb> elements whose <divn> context is unclear or ambiguous may be included in the immediately preceding <divn>, but not within any of its constituents. 
- End tags should always be used, even when they are marked as optional in the Document Type Definition (DTD). 

3. Transcription and Character Encoding

- "Electronic text at all levels of encoding should begin with the transcription of the first word on the first leaf of the original work. It may be impractical or undesirable to transcribe and encode certain features of the text, such as publisher's advertisements or indexes, but if at all possible, they should be included as links to page images. Any omissions of material found in the original work should be noted in the <editorialDecl> in the TEI Header." Annotations, marginalia, or other handwritten material in the source text should generally not be transcribed, except as may be required to meet the goals of the project. 
- Western European characters should be encoded as single characters from the ISO 8859-1 character set. Non-8859-1 characters should be encoded as character entity references rather than as 8-bit characters. Such characters must not be encoded as numeric entity references.  Character entity names should be drawn from standard entity sets such as:
 - ISO Added Latin 1
 - ISO Numeric and Special Graphic
 - ISO Publishing

- TEI Extra Classical Greek Letters
- ISO Greek Letters
- ISO Monotoniko Greek
- ISO Greek Symbols
- ISO Hebrew Letters

A chart comparing Greek character entity sets is available. 

- Punctuation and quotation characters may be encoded either as ASCII characters (where possible: , . : ; " ' ! & -) or as character entity references.  If character entities are used, they must if possible be chosen from standard entity sets such as ISO Numeric and Special Graphic or ISO Publishing. 
- Dashes longer than em dashes may be represented by a sequence of em dash character entities. 
- Hyphens should be retained, except for those dividing a word at the end of a print line.  However, even line-ending hyphens should be retained if the word normally contains a hyphen at that point. 

4. The TEI Header

Bibliographic descriptions of locally-created text collections should follow the same general guidelines regardless of the level of encoding of the source texts. Any differences based on encoding level will be noted where applicable.

- In local practice, all texts marked up at the Reading Level or higher are considered to be new editions of the work, and metadata decisions should be made accordingly.
- Throughout the header, all names (except those in prose paragraphs) should use either an established Library of Congress Name Authority File (LCNAF) form or a form consistent with LCNAF practice (for most languages last name, first name, middle initial). 

4.1 File description

Throughout the <fileDesc>, the type attribute of any <title> element must use the field tag of the MARC title field to which it corresponds. 

4.1.1 The <titleStmt>

- If the electronic work corresponds fully with the source text, the <title type="245"> in the <fileDesc><titleStmt> should be an exact copy of the 245 title in the <sourceDesc>.  If, however, the electronic text represents only a portion of the source, the <title type="245"> of the electronic text should be the title of that portion only.  In either case, the following text must be appended to the title:
 - : electronic text

for texts marked up at the Reading Level or higher, and

- o : electronic facsimile

for e-facsimile collections. The main title *must* also include an appropriately encoded level attribute. The number of nonfiling characters (if any) at the start of the title *must* be indicated in an appropriately encoded rend attribute.

- The <fileDesc><titleStmt> should repeat all <title> elements from the <sourceDesc> which apply to the electronic edition. Accordingly, uniform titles should occur in both places, while spine or cover titles should be entered only in the <sourceDesc>.
- The <author> element should not use a type attribute, and should not contain any <name> elements.
- The <author> element should be the same in both the <fileDesc><titleStmt> and the <sourceDesc>.
- The following roles in the creation of the *electronic* text must be recorded, when applicable, in the <fileDesc><titleStmt>:
 - o Author (using <author>)
 - o Principal researcher (using <principal>)
 - o Editor (using <editor>)
 - o TEI Markup (using a <respStmt> with <resp>TEI Markup</resp>)
 - o Funder (using <funder>), when creation of the electronic text has been funded by an external agency.
 - o Translator (using a <respStmt> with <resp>Translator</resp>), when the electronic text is a translation of the source work.
- Some roles in the creation of the *electronic* text may be recorded, if desired:
 - o Submitter (using a <respStmt> with <resp>Submitter</resp>)
 - o Technical advisor (using a <respStmt> with <resp>Technical advisor</resp>)
- Other roles (e.g., scanner, proofreader, typist) should not be recorded in the TEI Header.
- If one person has multiple roles (e.g., principal, editor, markup), those roles should be encoded separately using the appropriate elements.
- Responsibility statements (<respStmt>) should be constructed according to the following rules:
 - o The role <resp> should be recorded first, followed by one or more <name> elements.
 - o <resp> elements should consist only of the name of a role, without any connecting text.
 - o If individuals can be associated with specific roles, <name> elements should contain only personal names, and not those of organizational units.

4.1.2 The <editionStmt>

- An `<editionStmt>` must be present, containing one of the following forms of `<edition>`:
 - `<edition n="1">UW-Madison TEI edition</edition>`

for the first (local) edition, and

- `<edition n="2">Second UW-Madison TEI edition</edition>`

for subsequent editions. 

- Responsibility statements (`<respStmt>`) should be used only in subsequent editions, to record roles (such as editor) specific to those editions. All responsibilities for the first electronic edition will be recorded in the `<titleStmt>`. 

4.1.3 *The <extent>*

- The extent of the text file(s) should be recorded, using the most appropriate units, and the number of text files should be included. Sizes must be rendered as approximate: `<extent>ca. 250 Kb, in 2 files</extent>`.
- Files external to the text, but associated with it in some way (e.g., image files used for figures), may be recorded in the extent: `<extent>ca. 250 Kb, in 2 files: figures ca. 4.5 Mb, in 173 files</extent>`

4.1.4 *The <publicationStmt>*

- The following forms of publisher name and publication place should be used:

```
<publisher>University of Wisconsin-Madison Libraries</publisher>
<pubplace>Madison, Wisconsin</pubplace>
```

- Every electronic work must have a local identifier unique within the scope of the libraries' digital collections. The local Handles implementation will prevent name collisions, and a naming convention should be developed to facilitate the creation of identifiers. The identifier must be contained in an `<idno>` element, and its value must begin with the parent collection's ID followed by a period: `<idno type="Issue-ID">[collection.issue]</idno>`. 
- Any standard identifiers of the electronic work may be recorded in `<idno>` elements, using the following values for the `type` attribute: 
 - Number [default]
 - Issue-ID
 - Issue-Printed
 - SICI
 - URN
 - DOI
 - Handle
 - ISBN
 - ISSN

- Information regarding intellectual property ownership, copyright status, and access rights, when known, must be provided in an `<availability>` element. The University of Wisconsin Digital Collections Center (UWDCC) should be consulted regarding the proper wording to use for a particular collection. Access URLs may also be recorded as prose paragraphs in this element. 
- The `<date>` element should be used to indicate the date of publication of the electronic work. The element's content should be in prose format, but a `value` attribute must be supplied in the YYYY-MM-DD format defined by the W3C's proposed profile of ISO 8601. 

4.1.5 The `<seriesStmt>`

- If the electronic work is a component of a larger *electronic* collection, subcollection, or series, a `<seriesStmt>` should be created with the series type, title level, nonfiling title characters, and local identifier encoded in the appropriate attributes. 

4.1.6 The `<notesStmt>`

- If an abstract is created for a work (and is therefore not part of the source text), it should be encoded in the TEI header as a `<note>` with a `type` attribute of "Abstract". 

4.1.7 The `<sourceDesc>`

- All bibliographic information for the source text should be copied from the corresponding cataloging record whenever possible. If the cataloging is incomplete or nonexistent, the encoder should contact Central Technical Services for help in creating a bibliographic description of the source.  The number of nonfiling characters (if any) at the start of the series title *must* be indicated in an appropriately encoded `rend` attribute.
- If the electronic work is based upon an existing source (electronic or print), the `<sourceDesc>` must use a `<biblFull>` and its subelements to encode source text metadata. 
- If the text is originally created in digital form, the `<sourceDesc>` should consist of the following paragraph: 
 - `<p>Created in electronic form; no other source.</p>`
- The `<sourceDesc>` should reproduce (using `<title>` elements) all title fields found in the MARC catalog record for the source item. The only subfields which should be transcribed are |a (title), |b (remainder of title), |n (number of part), and |p (name of part). 
- Edition and series information, if available, should be encoded in `<editionStmt>` and `<seriesStmt>`, respectively.  An ID value (starting with the parent Collection ID) should be created for the series title, and a title level must also be

- encoded. Numbering within a series should be encoded in an `<idno>` element (with a type of "Issue-Printed") in the `<publicationStmt>`. The number of nonfiling characters (if any) at the start of the series title *must* be indicated in an appropriately encoded `rend` attribute.
- Notes describing the source text, edition, or variations should be included in a `<notesStmt>`.

4.2 Encoding Description

4.2.1 The `<projectDesc>`

- A prose description of the project may be included in the `<projectDesc>`.

4.2.2 The `<editorialDecl>`

- The encoding level, as described in this (or any other) document, should be recorded in the `<editorialDecl>`, citing the title, version, and URL of the document. A suitable `ID` attribute must be supplied for this statement. If the document is included in a collection also containing E-Facsimile materials, this statement (including `ID` value) **must** be provided. Any deviation from the source guidelines should be noted.
- Whenever the structure or sequence of a source text has been modified for some reason, an entry should be made in the `<editorialDecl>` noting the location (and possibly the reason) of the change, except when the markup (e.g., `<corr>`) makes it explicit.
- Decisions regarding correction, normalization, hyphenation, and retention of quotes in the source text should be noted in the `<editorialDecl>`. The following editorial policies should be used for collections at the Reading Level:

```
<p>No correction or normalization has been made to the source
text.</p>
<p>All 'soft' end-of-line hyphenation has been removed: any
remaining end-of-line hyphens are those considered to be part of
the original content, as opposed to occurring due to the vagaries
of the length of typographic lines.</p>
<p>Quotation and all other punctuation marks have been retained
in the transcription.</p>
<p>No segmentation or interpretation is provided with the
text.</p>
```

Collections encoded at higher levels are free to define other policies, as long as those decisions are recorded in the `<editorialDecl>`.

- Decisions regarding normalization and treatment of abbreviations or brevigraphs should be made by the editor or collection manager and be recorded in the `<editorialDecl>`.

4.2.3 The `<tagsDecl>`

- A Perl program (tagusage.pl) which will generate a <tagsDecl> element is available on the test server. This may be useful for document analysis prior to interface development. Contact the Library Technology Group for access information.

4.2.4 The <classDecl>

- A <taxonomy> must be declared (and an ID attribute created) for every keyword or classification scheme used in the <textClass>. These should be encoded as <bibl> elements containing the title of the scheme. 

4.3 Profile description

4.3.1 The <langUsage>

- If lang attributes are not used in the text, a <langUsage> element should be created with a paragraph stating the predominant language of the text.  If lang attributes are used in the text, a <langUsage> element must be created listing all languages referenced in the markup. The ID attribute of each <language> element must consist of a three letter language abbreviation defined in ISO 639-2, and the content should be the name of the language in English. 

4.3.2 The <textClass>

- If subject terms or other keywords are used for a text, they should be drawn from some recognized source such as Library of Congress Subject Headings [LCSH] or a domain-specific thesaurus such as the Thesaurus of Graphic Materials [TGM] or The Art & Architecture Thesaurus [AAT]. For each such scheme used, a separate <keywords> element must be created referencing (in the scheme attribute) a taxonomy declared in the <classDecl>. 
- Any locally-defined keywords must be created or approved by the collection manager. These keywords will not use a scheme attribute.
- The syntax of subject terms and subdivisions must follow the conventions of the scheme being used (final periods are optional). 
- For each keyword scheme, individual subject terms must be encoded as <term> elements. 
- If standard classification codes are used for a text, they should be drawn from some recognized source such as the Library of Congress Classification [LCC] or Dewey Decimal Classification [Dewey]. For each scheme used, a separate <classCode> element must be created referencing (in the scheme attribute) a taxonomy declared in the <classDecl>. 
- For each classification scheme, the classification code will be entered directly into the <classCode>; no sub-elements should be used. 

4.4 Revision Description

- Changes made during production of a text (*i.e.*, before initial publication) should be summarized in some convenient way in the `<revisionDesc>`.
- Changes made *after* the publication date must be itemized.
- The `<date>` of each `<change>` must be entered in the YYYY-MM-DD format defined by the W3C's proposed profile of ISO 8601. For summarized entries, date ranges can be indicated in ISO 8601 format by separating the start and end dates with a forward slash (/).

5 Levels of Encoding

- 5.1 E-Facsimile Level
- 5.2 Reading Level
- 5.3 Pedagogical Level
- 5.4 Scholarly Level
- 5.5 Dictionaries

5.1 E-Facsimile Level

This level is intended to support e-facsimile collections with no electronic text beyond OCR and metadata. The markup used for these collections is automatically generated, and is based on the data model developed for those collections. That specification is not within the scope of this document.

5.2 Reading Level

Texts at this level need to be marked up at a level minimally sufficient to support basic reading, browsing, retrieval, and navigation. The elements and attributes used at this level are, therefore, only those needed to format a text coherently on a screen.

Only the TEI elements specifically mentioned in these guidelines may be used in Reading Level texts. These elements are a subset of those found in the TEI Lite DTD.

5.2.1 Textual Divisions

- The predominant language of a work may be encoded as a `lang` attribute of the outermost `<text>` element containing the work. If it is used, the value of the `lang` attribute must be a three letter language abbreviation defined in ISO 639-2, and must correspond to the `id` attribute of a `<language>` element in the TEI Header.
- In order to meet campus accessibility policy [UWADA], any changes in the predominant language of a text must be marked using the global `lang` attribute. Where existing elements are available, add the `lang` attribute to the outermost element containing all and only the foreign language text. Where no appropriate element exists, use the `<foreign>` element and its `lang` attribute.

The value of the `lang` attribute must be a three letter language abbreviation defined in ISO 639-2, and must correspond to the `id` attribute of a `<language>` element in the TEI Header.

- The elements `<front>` and `<back>` should be used if and only if front and back matter are present in a text. 
- A `<titlePage>` element must be used to encode the Title Page. The following subelements must be used to encode its contents: 
 - `<docAuthor>` to encode the author's name. Enclose within a `<byline>` if there is additional text surrounding the author's name.
 - `<docTitle>` to encode the title, consisting of one or more `<titlePart>` elements with `type` attributes of:
 - `main` for the main title
 - `sub` for subtitle(s)
 - `<docImprint>` for publishing information, consisting of the elements `<pubPlace>`, `<docDate>`, and `<publisher>` as needed.
 - `<docEdition>` to contain text describing the current edition.
- `<div0>` elements should take one of the following values for the `type` attribute: 
 - Part [default]
 - Volume
 - Book

Please consult Section 2 for guidance on the proper use of `<div0>` elements.

- `<div1>` elements should take one of the following values for the `type` attribute: 
 - Section [default]
 - Abstract
 - Acknowledgements
 - Act
 - Appendix
 - Article
 - Bibliography
 - Chapter
 - Colophon
 - Contents
 - Cover
 - Dedication
 - Editorial
 - Errata
 - Foreword
 - Frontispiece
 - Glossary
 - Imprimatur
 - Index
 - Introduction
 - Lesson
 - Letter

- Masthead
- Notes
- Preface
- Scene
- Work

Please consult Section 2 for guidance on the proper use of `<div1>` elements.

- Content lists (*e.g.*, of illustrations) found in the front matter should use a `type` attribute of "Contents". 
- Individual works in a collection or anthology should be encoded as `<div1 type="Work">...</div1>`.
- `<div2>` elements should take one of the following values for the `type` attribute: 
 - Subsection [default]
 - Note
 - Frontispiece
 - Contents
 - Masthead
 - Foreword
 - Preface
 - Dedication
 - Abstract
 - Introduction
 - Imprimatur
 - Acknowledgements
 - Errata
 - Chapter
 - Article
 - Editorial
 - Work
 - Act
 - Scene
 - Letter
 - Notes
 - Index
 - Appendix
 - Glossary
 - Bibliography
 - Colophon
 - Cover
- The elements `<div3>` - `<div7>` may be used as needed for further subdivisions. The `type` attribute values for `<div2>` are available for all lower subdivisions. 

5.2.2 Headings and Closings

- All division-level heading information in a text should be encoded. If the text of the heading (or its translation into English) matches an item in the lists of accepted `<divn>` `type` attributes (*e.g.*, "Preface", "Glossary"), encode the heading (in English) in the `<divn>` `type` attribute. If the matching text is followed

by numbering (*e.g.*, "Chapter XV"), encode the numbering verbatim in the `<divn> n` attribute. In all cases, transcribe any heading text in a `<head>` element.

- If `<div2>` and lower subdivisions are analyzed and encoded, their headings should be encoded as for `<div1>` headings.
- The `<head>` element does not need to specify a `type` attribute. If the `type` attribute is used, it must take one of the following values:
 - `main`
 - `sub`
 - `duplicate`
- The `<epigraph>`, `<argument>`, and `<byline>` elements may be used as needed for prefatory matter.
- The `<trailer>` element may be used to encode a closing title or footer appearing at the end of a division of a text.

5.2.3 Verse and Drama

- Passages of verse typeset as verse will be encoded with `<1>...</1>` elements enclosed within a `<lg>` element for each typographically distinct group of one or more lines. The default `type` attribute value for `<lg>` will be "group"; other types of line groups should not be distinguished at this level except as noted below. Sequences of two or more line groups (`<lg>`) functioning as an integral unit (*e.g.*, a poem or section of a poem) should be contained within a single `<lg>` element.

Because the `type` attribute has the default specification `#CURRENT` in the DTD, if no value is supplied in the element occurrence the last specified value will be used. Therefore, only the first occurrence of the `<lg>` element must specify a `type` attribute if all subsequent occurrences are to have the same value.

- Where it is clear that a group of poetic lines constitutes a full poem (*e.g.*, by the presence of a title), the outermost `<lg>` element should encode the `type` as "poem"; otherwise the `type` will use the default value "group". If `type="poem"` is used for a line group, specify the `type` of subsequent line groups so they don't inherit the value "poem".
- When a passage of verse is embedded in a paragraph or similar element, it may be necessary to encode it as a quotation.
- Lines of poetry printed inline with surrounding prose text will not be marked up as poetry.
- Acts and scenes in dramatic texts should be encoded as `<divn>` elements with the `type` attributes "Act" and "Scene", respectively.
- A cast list in a dramatic work should be encoded with a `<list>` element, with each cast entry constituting an `<item>`.

- Speeches, speakers, and stage directions in dramatic works should be encoded with `<sp>`, `<speaker>`, and `<stage>`, respectively. 🐞

5.2.4 Lists and Tables

- Lists occurring in running prose text, even if labeled (*e.g.* "...three choices: 1) apples, 2) oranges, 3) bananas."), should not be marked up. 🐞
- Lists formatted as such (*e.g.*, each item on a separate line) should be encoded with `<list>` and `<item>` elements. The following types of list should be distinguished: 🐞
 - simple [default]
- Tables of contents, lists of illustrations, etc., should be encoded as `<list>` elements. 🐞
- Labels for list items must be transcribed when present in the source text. Simple labels such as numerals or letters should be encoded in the `<item>`'s `n` attribute. 🐞 For more complex labels it may be preferable to use the `<label>` element instead. 🐞
- There will be situations in which it is unclear whether a structure should be encoded as a list or as a table. In general, if the structure is formatted in multiple columns for which the columnar alignment is critical, encode it as a `<table>`. 🐞 Otherwise, encode it as a `<list>`. 🐞 Tables of contents are generally an exception to this rule.
- In some cases, it may be advisable not to encode a particularly complex table, but rather to include an image of the table as a `<figure>`. 🐞

5.2.5 Notes

- Notes occurring in running prose text (*e.g.*, set off with square brackets) should not be marked up. 🐞
- Footnotes should be marked up at their point of attachment to the text. While it is recommended that non-numeric note identifiers (*e.g.*, an asterisk) *not* be transcribed, it may be necessary in some texts to record the identifier in the `n` attribute of the `<note>` element. 🐞 If the text of a note spans several pages in the source, reassemble it into a single `<note>` element. 🐞
- Marginal notes should be marked up at their point of attachment to the text, if it can be determined unambiguously. Otherwise, place the `<note>` at the beginning of the paragraph (or comparable element) next to which it appears in the source text.
- Notes encoded at their point of attachment to the text must use one of the following values for the `place` attribute, according to where the note occurs in the source text: 🐞
 - foot [default]
 - margin

- o interlineal

- When end notes are gathered in a structural division of the source text, the section should be encoded as `<div1 type="Notes">` (for volume-end notes) or `<div2 type="Notes">` (for chapter-end notes). Individual notes in the section should be encoded with the `<note>` element, though in some cases (*e.g.*, appendices which are printed as notes) it will be preferable to encode them as `<div2 type="Note">`. It is not necessary to use the `place` attribute for end notes.
- It is not required that end notes be explicitly linked to their references in the text. Resources permitting, however, links should be created between each note and its reference using a `<ref>` element in the text and an `id` attribute on the `<note>` element, rather than using the `target` attribute on the `<note>` as described in the TEI Guidelines.
- Whether end notes are linked or not, their identifiers should be transcribed both at the reference point (if linked, as the content of the `<ref>` element) and in the note itself.

5.2.6 Figures

- The `<figure>` element must be used for all graphical content other than incidental decorations.
- For accessibility purposes, a `<figDesc>` element must be included for all figures, giving some indication of the content of the figure.
- All captions should be transcribed using `<head>` and `<p>` elements as appropriate. Other text in a figure may be similarly transcribed and encoded according to the structure of the text and the needs of the collections.
- Figures should be encoded at the place they occur in the text, if it can be determined unambiguously. Otherwise, put the `<figure>` at the beginning of the paragraph (or comparable element) next to which it appears.
- The `<figure>` element must contain an `entity` attribute corresponding to a definition for the reference image in the entity definition file. The interface software will locate other image resolutions based on existing directory and file naming conventions.
- The `<figure>` element's `rend` attribute value will determine the method used by the interface to display images. The following values are available:
 - o `thumb` [a thumbnail image will be linked to a higher-resolution image]
 - o `mmbib` [a reference image will be linked to a multimedia database record]
 - o `page` [an icon will be linked to an image of a page]
 - o `suppress` [no image will be displayed]

If the `rend` attribute is not present, a single, unlinked reference image will be displayed. If the 'mmbib' display is desired, the `<figure>`'s `id` attribute must equal that of the multimedia database record.

5.2.7 Quotations

- Passages of quoted text should be encoded (with the `<q>` element) only when set off from surrounding text with line breaks. 
- All quote marks in the source document should be retained. 

5.2.8 Letters

- Letters should be encoded as `<divn>` elements with a `type` attribute of "Letter". When the letter is quoted, this requires that the `<divn>` element be contained in `<text>` and `<body>` elements. 
- If the dateline is formatted on a separate line, encode it with `<dateline>...</dateline>`. 
- If the salutation and signature are formatted on separate lines, encode them with `<salute>...</salute>` and `<signed>...</signed>`, respectively. If either element spans more than one line, use a single element instance with `<lb>` elements indicating any line breaks. Do not enclose `<salute>` or `<signed>` within `<opener>` or `<closer>` elements unless they are already available (*e.g.*, by being required for a `<dateline>` element). 

5.2.9 Highlighting

- Text clearly distinguished from its immediate context (*e.g.*, in italics) should be tagged with the `<hi>` element. Use of one of the following `rend` attribute values (derived from XSL properties) is required: 
 - italic
 - bold
 - bolder
 - lighter
 - small-caps
 - uppercase
 - underline
 - overline
 - line-through
 - sub
 - super
 - larger
 - smaller
 - wider
 - narrower

Changes in typography attributable to structural function (*e.g.* bolding or capitalization of headings) should not be marked up.  If capitalized passages are transcribed in uppercase characters, they do not need to be marked with `<hi>` elements or `rend` attributes.

5.2.10 Reference and Linking

- Internal or external cross references should not be used at the Reading Level, except as suggested above for linking end notes.
- ID attributes should not be created unless automatically generated by software, or created for the purpose of linking end notes.

5.2.11 Errors and Correction

- Do not mark apparent errors in the source text, even if they seem to be obvious. 🗨️
- Clear errors in page numbering should be corrected (in a <pb> element) and noted in the <editorialDesc>.
- Missing page numbers should be supplied if they can be unambiguously determined from the immediate context. 🗨️ However, page numbers for unnumbered pages occurring at the beginning or end of a volume should not be supplied.

5.3 Pedagogical Level

Markup at this level should encode structure sufficient to enable search, retrieval, and display for the purposes of teaching or basic research in support of a subject or discipline. It may also provide for additional retrieval and display options as needed for purposes of text analysis. Additionally, it may contain references to external documents for purposes of text comparison, alignment, and reference.

All elements and attributes provided by TEILite may be used for markup at this level; in addition, some selected (limited) elements and attributes provided by the full TEI P3 (or its successors) may be used.

Guidelines for this level are forthcoming.

5.3.1 Textual Divisions

- All <div2> and lower subdivisions should be analyzed and encoded, and their headings should be encoded as for <div1> heads.
- Anthologies or collections should use a <group> element to enclose <text> elements comprising individual works.

5.3.2 Headings and Closings

5.3.3 Verse and Drama

- Line groups should use one of the following values for the `type` attribute:
 - `group` [default]
 - `poem`
 - `stanza`
 - `verse`
 - `strophe`
 -

-
- Lines of poetry printed inline with surrounding prose text should be marked up with appropriate verse elements.

5.3.4 Lists and Tables

5.3.5 Notes

- Notes encoded at their point of attachment to the text must use one of the following values for the `place` attribute, according to where the note occurs in the source text:
 - `foot`
 - `margin`
 - `interlineal`
 - `inline`

5.3.6 Figures

5.3.7 Quotations

5.3.8 Letters

5.3.9 Highlighting

5.3.10 Reference and Linking

- In order for cross references to work in our current environment, all `<ref>` and `<ptr>` elements must include a `targetType` attribute indicating the element type to which they are pointing.
- The current infrastructure only recognizes a single `ID` value as the value of a `target` attribute; one-to-many links are not supported.
- References encoded with `<xref>` or `<xptr>` elements are not currently supported.
- [Method for constructing `id` attributes]

5.3.11 Errors and Correction

5.4 Scholarly Level

The markup used for this level will be determined according to the needs of the researcher. Hence, no specific guidelines will be developed for Scholarly Level collections. Still, to the extent possible scholars should make every effort to follow the various guidelines defined for the other levels. Doing so will simplify the task of indexing and interface development and should also provide users with a more consistent environment.

5.4.1 Entry structure

- Every entry must have an `id` attribute, preferably derived from the lemma.
- Homographic forms should be encoded in separate `<entry>` elements, distinguished by means of `n` attributes.
- If it is desired that editorial or authorial responsibility be recorded at the level of the entry, the `<respons>` element (from the tag set for Certainty and Responsibility) may be used for this purpose.

5.4.2 Headwords

- All forms of the lemma must be enclosed in a single `<form>` element with a `type` attribute of "lemma". The `lang` attribute must also be used to encode the language of the headword.
- If syllabic, morphological, or other boundaries are indicated in the headword, this form must be encoded in an `<orth>` element with a `type` attribute of "marked". An unmarked form of the headword must also be included, as an `<orth>` element with a `type` attribute of "std". The standard form may, however, be created automatically during preprocessing.
- Pronunciation [forthcoming]

5.4.3 Grammatical Information

- Part of speech and other grammatical categorization, if included, must be encoded as a `<gramGrp>` element containing a `<gram>` element for each grammatical category. Each `<gram>` element must contain a `type` attribute with one of the following values:
 - `asp` (aspect)
 - `case`
 - `dgr` (degree)
 - `fragment`
 - `gen` (gender)
 - `itype` (inflectional type)
 - `mood`
 - `num` (number)
 - `per` (person)
 - `pos` (part of speech)
 - `subc` (subcategorization/case governance)
 - `tns` (tense)
 - `voice`
- `<gramGrp>` elements may also contain an `opt` attribute with a value of `y` or `n`. The content of `<gram>` may be coded, abbreviated, or spelled in full, and will be displayed as entered.

5.4.4 Morphological forms

- Examples of inflected forms of the headword must be grouped in a <form> element with a type attribute of "infl-grp", even if there is only one such form. There may, however, be multiple inflectional groups, so the <form> element may be repeated. Individual forms, whether abbreviated (e.g., suffix only) or full, must be encoded as <orth> elements with an extent attribute indicating the extent of the form. Values include:
 - full
 - phrase
 - suff (suffix)

The interface also supports use of a rend attribute with the value "parens".

- If grammatical information is included with the inflected form (e.g., case or tense), it must be encoded as a <gram> element with an appropriate type attribute as described above. This <gram> and the associated <orth> must be contained within a <form> element with a type attribute of "inflected". There may be one or more such <form> elements within a <form type="infl-grp">.

5.4.5 Cross References

5.4.6 Definitions and Translations

5.4.6.1 Collocations or idiomatic phrases

5.4.6.2 Usage hints

5.5 Dictionaries

6. References

- [AACR2] Joint Steering Committee for Revision of AACR. *Anglo-American cataloguing rules, second edition* Chicago: American Library Association, 1986. See especially ch. 22, "Headings for Persons".
- [AAT] Getty Research Institute. *Art & Architecture Thesaurus Introductory Page (Getty Vocabulary Program)* <URL: http://shiva.pub.getty.edu/aat_browser/>
- [Dewey] *Dewey Decimal Classification Home Page (OCLC Forest Press)* <URL: <http://www.oclc.org/oclc/fp/>>
- [ISO639] Library of Congress. *ISO 639-2 Registration Authority - Library of Congress* <URL: <http://lcweb.loc.gov/standards/iso639-2/>>
- [LCC] Library of Congress. Cataloging Policy And Support Office. *Library of Congress Classification Outline* <URL: <http://lcweb.loc.gov/catdir/cpso/lcco/lcco.html>>
- [LCSH] Library of Congress. *Library of Congress subject headings* Washington: Library of Congress, 1975-

- [TEIP4] Burnard, Lou, and C. M. Sperberg-McQueen, eds. *Guidelines for Electronic Text Encoding and Interchange* <URL: <http://www.tei-c.org.uk/P4X/>>
- [TEIU5] Burnard, Lou, and C. M. Sperberg-McQueen. *TEI U5: Encoding for Interchange: an introduction to the TEI* <URL: <http://www.tei-c.org.uk/Lite/>>
- [TEIXML1] TEI and XML in Digital Libraries Workshop. *TEI Text Encoding in Libraries* <URL: <http://www.indiana.edu/~letrs/tei/>>
- [TEIXML2] TEI and XML in Digital Libraries Workshop. *TEI/MARC 'Best Practices'* <URL: <http://www-personal.umich.edu/~jaheim/teiguide.html>>
- [TGM] Library of Congress. Prints and Photographs Division. *LC Thesaurus for Graphic Materials I: Subject Terms (TGM I)* <URL: <http://lcweb.loc.gov/rr/print/tgm1/>>
- [UWADA] University of Wisconsin-Madison. Office of the ADA Coordinator. *UW-Madison: Policies: Affirmative Action and Equal Opportunity*
- [UWEFacs] Gorman, Peter C., ed. *UW-Madison Digital Library Data Dictionary: Electronic Facsimile Collections* <URL: <http://uwdcc.library.wisc.edu/resources/efacs/EFacsDataDictionary.shtml>>
- [UWGreek] Gorman, Peter C. *Display of Greek characters: UW-Madison Libraries* <URL: <http://libtext.library.wisc.edu/oticons/local/greek/>>
- [W3C-DATE] World Wide Web Consortium. *Date and Time Formats* <URL: <http://www.w3.org/TR/NOTE-datetime>>
- [XSL] World Wide Web Consortium. *Extensible Stylesheet Language (XSL)* <URL: <http://www.w3.org/TR/xsl/>>

The TEI Markup Guidelines Working Group:

- Steve Dast
- Edie Dixon
- Peter Gorman (chair)
- Luis Villar
- Barbara Walden
- with assistance from Jamie Woods