A TEI Project

Best Practices for TEI in Libraries

Author: TEI SIG on Libraries

Editors: Kevin Hawkins, Michelle Dalmau, and Syd Bauman

2011-10

Version 3.0 (October 2011)

Table of contents

1. Introduction

This document is the third version of a document formerly known as *TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices*, which has been updated to comply with the Text Encoding Initiative's Guidelines for Text Encoding and Interchange (P5). These guidelines are intended for use in large, library-based digitization projects, but may be useful in other scenarios as well. This version of the *Best Practices for TEI in Libraries* was created and is maintained by the *TEI in Libraries: Guidelines for Best Practices* Working Group.

There are many different library text digitization projects, serving a variety of purposes. With this in mind, these *Best Practices* are meant to be as inclusive as possible by specifying five encoding levels. These levels are meant to allow for a range of practice, from wholly automated text creation and encoding, to encoding that requires expert content knowledge, analysis, and editing. The encoding levels are not strictly cumulative: while higher levels tend to build upon lower levels by including more elements, higher levels are not supersets because some elements used at lower levels are not used at higher levels—often because more specific elements replace generic elements.

In brief, the encoding levels (with approximate examples) are as follows:

Level	Description	Example of encoding of **Alger Hiss document**	Display example
Level 1	The text is generated through OCR, is subordinate to the page image, and is not intended to stand alone as an electronic text (without page images). Encoding is done to assist in full text searching.	Alger Hiss document	example
Level 2	The text is generated through OCR and is mainly subordinate to the page image, though navigational markers (textual divisions, headings) are captured.	Alger Hiss document	example
Level 3 (example)	The text is created by conversion, either by way of OCR or keyboarding. Some structural elements of the text are encoded. The text may be used with or without page images.	Alger Hiss document	example
Level 4 (example)	The text is generated either through corrected OCR or keyboarding and is able to stand alone without page images in order for them to be read by students, scholars, and general readers.	Alger Hiss document	example

Level 5 (example)	The text is generated either through corrected OCR or keyboarding and is able to stand alone without page images, as in Level 4. In addition, the tagging requires substantial human intervention by encoders with subject knowledge.	(none)	example

In these *Best Practices*, use of elements and attributes tends toward explicitness for ease of processing even though a human or possibly machine reader might be able to make inferences based on context. Only those elements and attributes mentioned below are recommended for use in encoding based on these *Best Practices*; use of other TEI elements and attributes is not recommended. Consult the full TEI Guidelines for guidance on use of elements and attributes beyond what is described below.

As guidelines rather than a specification, these *Best Practices* use *should* instead of *must* in nearly all cases (except where a practice is required by P5), with optional practices indicated by *may*. However, to encourage conformance to these *Best Practices*, the ODD files will generate schemas that *require* use of recommended elements (indicated by *should*).

These *Best Practices* specify a recommended archival storage format. Local system needs may require transformation of documents in this archival format to another XML format for use by a local indexing or delivery software.

2. Relationship to TEI Tite

The TEI Tite customization of the TEI Guidelines was developed as a subset of the TEI to be used as a vendor specification for outsourced encoding of the type often initiated by libraries, archives and other cultural heritage organizations. This *Best Practices* document was created to support in-house encoding that adheres as closely as possible to common TEI practice and library standards yet still leaves room for variation in local practice.

If a library uses TEI Tite for outsourced encoding, it should find that converting files from the TEI Tite format to a format conforming to these *Best Practices* is not difficult. TEI Tite files may be converted to *Best Practices* Level 3 with some loss of granularity; or to Level 4 with some additional markup with minimal human intervention. (The reason Level 3 does not contain as many elements as TEI Tite is to allow for use of this encoding level, whether for encoding of born-digital source documents or for upgrading Level 1 or Level 2 texts, with less human intervention than would be required by TEI Tite.)

These *Best Practices* are meant to complement the TEI Tite customization of the TEI Guidelines. Whereas TEI Tite is meant for vendors who need exact specifications for encoding, these *Best Practices* document how a library or other large-scale encoding project might create conformant TEI documents out of vendor-generated or locally-created TEI documents. TEI Tite lacks header metadata and elements for encoding textual structures of possible interest to libraries; however, once Tite documents are transformed to a TEI-conformant encoding used by an institution, these *Best Practices* can serve as a point of reference for developing the TEI header and applying richer markup as reflected in Level 4 or 5 of these *Best Practices*.

For a comparison of the TEI Tite schema to these *Best Practices*, see TEI Tite's Appendix A.

3. General Recommendations

3.1. Standards and Local Practice

The goal of the TEI is interchange, not interoperability. While seamless interoperability of texts created for different purposes is an elusive goal, use of a common markup vocabulary and syntax greatly aids interchange. Nevertheless, keep in mind that others—even within your organization—may use your texts in the future for other uses than you intended in your encoding.

An encoding project should strive for internal consistency and for use of standards so that the data can be modified or enhanced in the future with ease. In cases where local practice deviates from standards, there should at least be internal consistency in the local practice.

3.2. Transcription

When reformatting to digital media using any level of encoding, the electronic text should begin with the transcription of the first word on the first leaf of the original work. At lower levels of encoding, it may be impractical or undesirable to transcribe and encode certain features of the text, such as publisher's advertisements or indexes; at Level 4 and above, the transcription should be complete. Any omissions of material found in the original work should be noted in the <editorialDecl> in the TEI header.

3.2.1. Punctuation

While the TEI gives the option of removing punctuation marks that could be encoded using XML tags, such as quotation marks or end-of-line hyphens, from the transcribed text, it is recommended to leave these as XML character data in order to simplify rendering of the encoded text as it appears in the source document. For example:

```
<p>She said, <said>"Nobody uses the term
 <soCalled>'electronic text'</soCalled> anymore"</said>!</p>
```

Use <editorialDecl> in the TEI header to record whether such punctuation has been removed.

When removing punctuation marks whose rendering in the source document is worth preserving, use the **@rend** attribute on the appropriate element to describe the punctuation marks that were removed. For example:

```
<p>She said, <said rend="quotes: '"' '"'">Nobody uses the term
 <soCalled rend="quotes: '‘' '’'">electronic text</soCalled> anymore</said>!</p>
```

3.2.2. Hyphenation

Encoding end-of-line, end-of-column, and end-of-page hyphenation varies considerably in the TEI community. Some capture all hyphens found on the printed page, while others remove those in the middle of words not normally hyphenated for easier implementation of full-text retrieval. If preserving hyphens, some will capture all hyphens using the same character, while others will distinguish hyphens that must be present in any case (often called hard hyphens) and those that are only present by virtue of being at the end of a line, column, or page (often called soft hyphens).
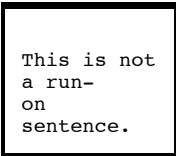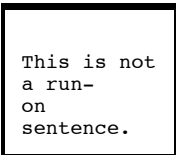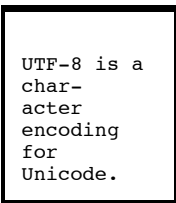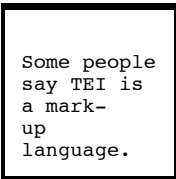
This issue is complicated by the fact that Unicode prescribes use of a soft hyphen not for a *visible hyphen that might have been absent* but instead for a place where a hyphen *might occur*. Furthermore, it includes a non-breaking hyphen, used in cases like 're-creation' (meaning to create again, as opposed to recreation, meaning relaxation), in addition to a regular hyphen, which would normally count as a word boundary. In short, Unicode is oriented toward electronic text that may be processed with a computer in various ways, not toward capturing source documents.

Since OCR software relies on dictionaries to determine the probability not simply of characters but of whole words, it is often able to capture hyphenation in different ways, per the needs of a specific project.

At Levels 1 and 2, do not attempt to disambiguate different uses of hyphens. Encode all hyphens appearing in the source document using character U+2010 if possible; alternatively, use the semantically ambiguous U+002D.

At Level 3, optionally distinguish uses of the hyphen with the **@break** and **@rend** attributes on <lb>, <cb>, and <pb> elements as appropriate. At Level 4, the use of the **@break** and **@rend** attributes on these elements is mandatory.

Here are some examples of use of these attribute on the <lb> element. (If these hyphens occur at a column or page break instead of a line break, use <cb> or <pb> instead.)

Colloquial name	Appearance in source document	Encoding	Note
Hard hyphen	This is not a run- on sentence.	`This is not a run-<lb break="no" rend="keep-hyphen"/>on sentence.`	The use of no as the value of the **@break** attribute indicates that the encoder considers "run-on" to be a single orthographic token (loosely speaking, a single word).
Hard hyphen	This is not a run- on sentence.	`This is not a run-<lb break="yes" rend="keep-hyphen"/>on sentence.`	The use of yes as the value of the **@break** attribute indicates that the encoder considers "run-on" to consist of two separate orthographic tokens.
Soft hyphen	UTF-8 is a char- acter encoding for Unicode.	`UTF-8 is a char-<lb break="no"/>acter encoding for Unicode.`	The use of no as the value of the **@break** attribute indicates that the encoder considers "character" to be a single orthographic token.
Unclear case	Some people say TEI is a mark- up language.	`Some people say TEI is a mark-<lb break="maybe"/>up language.`	The use of maybe as the value of the **@break** attribute indicates that the encoder is unsure whether "mark-up" is a single orthographic token.

Do not confuse the following characters with hyphens:

- en dash (U+2013)
- em dash (U+2014)
- minus sign (U+2212)

3.3. Filenames

A filename scheme that is internally consistent should be established for the project.

If it is likely that the files will need to be used on more primitive devices (MS-DOS computers or unextended ISO 9660 CDs) it may be useful to limit names to 8 characters (limited to the 26 lower case letters of ASCII, digits, hyphens, and underscore), a dot, and an extension of 3 alphanumeric characters. Likewise, if you will access files using a version of Apple Filing Protocol (AFP) before 3.0, filenames longer than 31 bytes are likely to be corrupted, so you may wish to limit filenames to 31 single-byte (e.g., ASCII) characters.

Otherwise, consider the following best practices when determining the file name scheme for your project:

- Each filename should contain an identifier that uniquely specifies a single digital object within the parent collection (e.g., a parent collection of text, images and other related materials)
- Each filename should be fully specified. It should not just be a sequence number that is dependent on location within a directory structure for context
- Filenames should not include spaces
- Filenames should follow predictable case constructions (e.g., all lowercase, camelCase, etc.)
- The first character of the filename should be an ASCII letter ('a' through 'z' or 'A' through 'Z') to comply with current restrictions on identifiers by many programming and XML-based metadata languages
- The base filename may include only ASCII letters ('a' through 'z' and 'A' through 'Z'), ASCII digits ('0' through '9'), hyphens, underscores, and periods. Refrain from using other characters and limit period usage to only once (to separate base name from file extensions).

3.4. URIs

A number of attributes take a URI (Uniform Resource Identifier) as their value. Note that in addition to the full form of reference defined by URI syntax, these attributes can take a relative reference (e.g., `filename.ext`) or a fragment identifier (e.g., `#foo`).

3.5. Textual Divisions

An encoding project should use only numbered divisions (i.e., <div1>, <div2>, etc.) or unnumbered divisions (i.e., <div>) but not both. This applies both within a TEI document (i.e., within <front>, <body>, <back>, even if nested within <group> or <floatingText>) and across TEI documents in any given collection. Keep in mind that numbering of textual subdivisions starts over (at <div1>) within <floatingText> nested inside a subdivision, so any software that expects to process nested numbered divisions within a document will need to account for this.

The choice of numbered or unnumbered divisions should be documented with the <tagUsage> element in the header. See 4.6, Element Recommendations for the TEI Header, below.

Whether numbered or unnumbered divisions are used, the **@type** attribute of the division element is not recommended at Level 1 (because only one encoded division in the text exists), is optional at Level 2 (because the division-level metadata need not classify these divisions), is recommended at Level 3 (for broad yet useful analysis of text divisions), and is strongly recommended at Levels 4 and 5 (for full analysis of the text structure).

3.6. Page Breaks

Page breaks should be encoded using the <pb> element, with the value of the **@n** attribute denoting the number of the page whose text *follows* this element. The <pb> element should always be contained within a text division for ease of retrieval with indexing software. For example, a page break that occurs between chapters 2 and 3 should be encoded right after the opening tag of the textual division that opens chapter 3 rather than before the closing tag of the division that ends chapter 2.

3.7. Linking Between Encoded Text and Images of Source Documents

There are three recommended mechanisms for linking between the encoded text and facsimile page images of source documents. Projects may use any of the following methods:

- Use the **@facs** attribute on a <pb> element to point to the corresponding page image using a URI.
- Use the <facsimile> element to define a set of images that corresponds to the text, in conjunction with the **@facs** attribute on a <pb> element to point to the corresponding page image using a URI.
- Use the **@xml:id** attribute on each <pb> element and a METS document to provide correspondence between <pb> elements and one or more facsimile page images (e.g., master, web derivatives, etc.).

For those projects relying on the Metadata Encoding and Transmission Standard (METS), the **@xml:id** attribute is used as a conceptual identifier for content as opposed to an explicit pointer (i.e., **@facs** attribute) to a specific representation of that content. These identifiers are then used to generate a METS document that bundles the various content types (e.g., master image files, derivative image files for Web delivery, PDFs, etc.), explicitly lists all versions of the content, and defines the relationships between the constituent parts. This is achieved through the use of the <mets:fileSec> and <mets:structMap> sections of the METS document (see sample METS document for a TEI project).

3.8. General Guidelines for Attribute Usage

These *Best Practices* provide recommended usage of attributes as used in the TEI header and within the body of the TEI document (within the <text> element), as evidenced by attributes used in encoding example snippets and the prose description of this document.

Scores of attributes are available for use within <text>, and a list of those recommended for use in Tite documents is included in Appendix A.

In this section is general advice on the use of particular attributes commonly needed for library encoding projects. (All of the attributes below are commonly used on various elements, but not every element requires or even allows these attributes.)

3.8.1. @type

Constructing a list of acceptable attribute values for the **@type** attribute for each element, on which everyone could agree, is impossible. Instead, it is recommended that projects describe the **@type** attribute values used in their texts in the project ODD file and that this list be made available to people using the texts. It is worth noting that, at present, Roma, the web front-end editor for ODD files, does not have a mechanism for providing this documentation — it should be added to the ODD file directly. For a list of standard names and definitions of bibliographic features of printed books, see *ABC for Book Collectors* by John Carter (8th edition, New Castle, Del. and London: Oak Knoll Books and the British Library, 2004, available online at http://www.ilab.org/images/abcforbookcollectors.pdf).

3.8.2. @n

This attribute is sometimes used to number elements for machine processing, but it often includes data represented in the source document, such as page numbers or footnote numbers. Example: pb n="456"/

3.8.3. key and ref

These attributes are both available on a variety of elements including <persName>, <orgName> <author>, and <title>. They are used to reference external metadata about the content of the element. The **@key** attribute may contain any string of Unicode characters, whereas the **@ref** attribute contains a URI (including a relative one, as discussed above). While **@key** may supply any identifier, there is no mechanism internal to XML for checking that the value of this attribute is valid.

For example,

```
<author>
 <persName type="marc100" key="lccn-n78-95332">Shakespeare, William, 1564-1616</persName>
</author>
```

gives a project-specific key (in this case lccn-n78-95332) for this name in the Library of Congress Name Authority File. Values of **@key** attributes may be partially explained in a non-machine-readable way through the use of a <taxonomy> element:

```
<taxonomy xml:id="lccn">
 <bibl>Library of Congress Control Number</bibl>
</taxonomy>
```

Alternatively, use **@ref** with a URI fragment identifier, corresponding to the value of **@xml:id** given elsewhere. For example, in the transcription of the text, use

```
<placeName ref="#tgn_7012924">Indianapolis</placeName>
```

which would be defined in a controlled vocabulary elsewhere:

```
<listPlace>
 <place xml:id="tgn_7012924">
  <placeName>
   <settlement type="city">Indianapolis</settlement>
   <region type="state">Indiana</region>
  </placeName>
 </place>
</listPlace>
```

Readily available software can then check when it encounters ref="#tgn_7012924" that xml:id="tgn_7012924" exists elsewhere in the document.

In general it is recommended to use **@ref** when the metadata object being referenced is accessible via a URI (e.g., is on the web), and **@key** when it is not. To avoid ambiguity in referencing external data sources, it is recommended not to use both attributes on the same instance of an element.

3.8.4. @rend and @rendition

At levels 3 and above, the **@rend** and **@rendition** attributes may be used when it is desirable to record information about how the textual feature was displayed in the source document.

Never use these attributes on header elements: metadata is transcribed and possibly regularized, as in a catalog record, but its exact appearance is not meant to be captured.

If a project is normalizing the rendering of text objects (for example, such that all titles should be italicized, regardless of how they appeared in the source document), there is no need to use these attributes; instead, a stylesheet will determine that all titles are displayed in italics.

However, if a project is faithfully recording the rendering in the source document, one of these attributes should be used to indicate this rendering, either on all elements to be rendered differently from the surrounding text or on all elements whose rendering does not follow the default stylesheet.

For the value of the **@rend** attribute, use only valid CSS properties and values as in a declaration-block but without curly braces. For example:

```
<foreign rend="font-style: italic">
```

```
<title rend="text-decoration: underline; font-size: x-large">
```

Alternatively, use the **@rendition** to give an internal scheme:

```
<foreign rendition="#i">
```

documented with the <rendition> in the header:

```
<rendition xml:id="i" scheme="css">font-style: italic</rendition>
```

Use of the **@rendition** attribute and element offers an additional level of indirection, decreasing the total number of keystrokes and possibly reducing the chance of typos being introduced in the encoding.

3.8.5. @xml:lang

Used to indicate the natural language of the content of an element. It is generally not used for children of the <text> element at Level 1 or Level 2 but is common at Level 3 and above. See the data.language datatype in the TEI Guidelines.

4. Structure of a TEI Document

Element	Description
TEI xml:id="___" xmlns="http://www.tei-c.org/ns/1.0"	The root element of a TEI document. Use of the **@xml:id** attribute is recommended, giving the same unique identifier for the TEI document as in `teiHeader/fileDesc/publicationStmt/idno`.
├ teiHeader xml:lang="___"	The <teiHeader> contains metadata about the TEI document. The **@xml:lang** is recommended; it indicates the language used for the metadata describing the document.
├ <facsimile>	The <facsimile> defines sets of images that correspond with the text. This element should only be used if page images are included and if this particular mechanism for linking page images is chosen. See between encoded text and images of source documents|Linking between encoded text and images of source documents.
└ The text xml:lang="___"	The <text> element contains the encoded transcription of the source document. The **@xml:lang** attribute is recommended; it indicates the primary language of the source document.

The child elements of the <teiHeader> and <text> elements are described below.

4.1. The TEI Header

4.1.1. Reference

- Chapter 2, TEI Header, P5 Guidelines

Note that this is not a complete customization. It is just one specification group that is used by each of the customizations for levels 1–4.

4.1.2. Introduction

The TEI header is a metadata record for an encoded text. It includes bibliographic information related to the electronic document and, if appropriate, the bibliographic data for the original analog source document from which the electronic edition was created. The TEI header often includes a description of the encoding decisions or practices used to create the electronic document. While TEI Lite calls the header 'the electronic title page', it actually more closely resembles a catalog record with additional data not routinely stored in MARC records.

As with any descriptive metadata, the metadata in the TEI header can serve multiple audiences. In the local context, a TEI header provides metadata about the TEI document, its source, and its provenance. The TEI header may be used for metadata exchange, to automatically create indexes (author lists, title lists) for a collection of TEI documents, and to aid in browsing heterogeneous TEI documents. TEI headers may also be used as a basis for other metadata records (such as MARC or Dublin Core), though generation of other formats may require human intervention because they often are more granular, or have different granularity, than TEI headers.

4.1.3. The TEI Header and MARC

While a TEI header is often perceived as similar to or at least related to a MARC record, a TEI header does not typically have a one-to-one correspondence with a MARC record. One TEI header may be described by multiple MARC analytic records, or one MARC record may be used to describe a collection of TEI documents with individual headers. Furthermore, while a MARC record captures metadata about a bibliographic entity in a library's collection, a TEI header records information both about an encoded text and about the source document for that encoded text. Each institution and even each project may have a different approach to the way electronic texts are created in TEI and then represented in a larger public catalog through MARC. At one institution, the same unit (e.g., a cataloging department) may be responsible for creating both TEI headers and MARC records, while at other institutions the work may be distributed among different units. Within the library domain, metadata or cataloging experts are usually required for at least review and standardization of both the TEI header and the MARC record. In order to allow automatic generation of TEI headers from MARC records and MARC records from TEI headers, some elements (like <author>) contain content not typical for TEI practice but necessary due to a lack of granularity in the MARC format.

4.1.4. The TEI Header and Other Metadata Schemas

Several other descriptive metadata schemas are prevalent within the library domain, including Dublin Core (DC), Dublin Core Qualified (DCQ), and the Metadata Object Description Schema (MODS). Each of these schemas contains elements that capture the same data as many of the elements in the TEI header. As with MARC, a variety of automated or manual workflows can be implemented to crosswalk metadata from one standard to another and provide for increased sharing of metadata about electronic texts in larger contexts. In particular, DC and MODS are common schemas used by the Open Archives Initiative (OAI) and may be particularly valuable for sharing metadata across institutions. Unfortunately, there is currently no mechanism for specifying that the content of an element should be drawn from an outside metadata source or that this outside metadata source should supplement the content of the element. In the absence of such mechanisms, users of these *Best Practices* may use the <idno> element to supply identifiers for outside metadata records and may supply identifiers for certain authority records using the **@key** or **@ref** attributes, allowed on certain elements.

4.1.5. Determining Data Values for the TEI Header

Within the library domain, there are several authoritative publications on how to create bibliographic and descriptive metadata for objects. These are usually called "content standards"; two prominent examples are the *Anglo-American Cataloging Rules Second Edition* (AACR2) and the *International Standard Bibliographic Description for Electronic Resources* (ISBD(ER)). These standards are extensive and outline a set of rules that enforce consistency across a voluminous amount of metadata. It is recommended that metadata about the source document included in the header be taken from the catalog record for the source document. However, there may be cases when this information is incomplete or insufficient. Furthermore, creation of other TEI header elements may require more context than is available simply from the encoded text. But the analog object may not be available, so the TEI header creator will need access to digitized images or other verifiable information to create accurate metadata. The following sources of information are recommended in creating the TEI header:

1. For an electronic document with a digitized title page and title page verso, the chief source of information is the information coded as the title page and title page verso. Use other sources of information from a physical source document if absolutely certain that it is the source.
2. If there is no digitized title page but the header creator knows the physical source document from which it was derived, the header creator should refer to that source document for metadata creation. Note that a lack of a title page may be for one of many reasons: for example, the original document is a manuscript item, or the electronic edition is a portion of the original object (a poem or short story that was published in a collection or an article from a serial). In all cases, it is recommended that important bibliographic evidence, such as a digitized image of the title page and title page verso for a collection, be provided to the header creator, even if just a piece of the collection is used.
3. If no title page is present and there is no evidence from a source document, the header creator may assign a title and author, if appropriate, enclosing the information in square brackets (the standard English-language convention for editorial interjections).

4.1.6. Element and Attribute Recommendations for the TEI Header

Below is documentation on use of elements and attributes within the <teiHeader> element. These recommendations apply to all levels of encoding. Gray boxes in the source document column indicate that while the corresponding TEI element describes the TEI document, the value of this field is often derived from metadata about the source document, to be found in the MARC fields listed.

Element	Description	Equivalent in MARC when cataloging the TEI document	Equivalent in MARC for the source document
teiHeader xml:lang="___"	The <teiHeader> contains metadata	040 $b	n/a

				about the TEI document. The **@xml:lang** attribute is recommended; it indicates the language used for the metadata describing the document.		
├		<fileDesc>		The <fileDesc> contains bibliographic metadata about the TEI document. One of its child elements, <sourceDesc>, describes the source document from which the TEI document was created.	n/a	n/a
│	├	<titleStmt>			n/a	n/a
│	│	├	title type="_"	One or more <title> elements are required to give the title of the TEI document being created. It is suggested that titles be constructed based on the source document according to a national cataloging code. Use of the **@level** attribute is not recommended since it does not apply to a TEI document in a collection. Use of the **@type** attribute is recommended. It should have one of the following values as suitable in local practice: • main • sub • alt • short • desc • translated • marc245a (used for the title proper and alternative title according to the national cataloging code) • filing (used for a version of the title with initial articles removed, to be used for sorting titles alphabetical • ly but not for display) marc245b (used for the the remainder of the title information — parallel titles, titles subsequent to the first, and other title information — according to the national cataloging code) • uniform (used for a uniform title according to the national cataloging code)	• 130 • 240 • 245 $a,$b • 246	• 130 • 240 • 245 $a,$b • 246

│	│	├	<author>	One or more <author> elements (one name per element) are used to encode the names of entities primarily responsible for the content of the TEI document—usually, the author(s) of the source document. Use <persName> or <orgName> when applicable. Whenever possible, establish or use the form of the name from a national name authority file. Examples:
				
				• `<author>`
				`<persName>Shakespeare,`
				`William, 1564-`
				`1616</persName>`
				`</author>`
					• 100
					• 110
					• 534 $a = 1st author
					• 700
					• 710
					• 711
						• 100
						• 110
						• 700
						• 710

```
<author>
<orgName>National
Organization for
Women</orgName>
</author>
```

```
<author><persName>X,
Malcolm</persName>
</author>
```

```
<author>
<persName>Thomas
(Anglo-Norman poet)
</persName></author>
```

```
<author>
<persName>Catherine
II, Empress of
Russia</persName>
</author>
```

```
<author>
<persName>Joannes,
Actuarius, 13th/14th
cent.</persName>
</author>
```

| | | | ├ | <editor> | If applicable, use one or more <editor> elements (one name per element) to encode the names of entities besides those in <author> elements that acted as editors of the TEI document— usually, the editor(s) of the source document. If considered appropriate by the encoding project, the editor of the TEI document should be entered here. Use <persName> or <orgName> when applicable. Whenever possible, establish or use the form of the name from a national name authority file. Unlike in the TEI Guidelines, do not use this element for translators, illustrators, compilers, or other roles not generally considered an editor. Therefore, do not use the **@role** attribute. | • 700
• 710 | • 700
• 710 |
| | | | ├ | <respStmt> | Record the names of other persons or organizations, one responsibility or party per <respStmt>, that have responsibility for the intellectual or artistic content of the TEI document— often by transitivity from the source document—not covered by <author> and <editor>. This includes translators, illustrators, compilers, proofreaders, encoders, and those who wrote a preface or introduction. Each <respStmt> should contain either:

    • one <resp> followed by one or more of <persName> or | • 500
• 700
• 710 | • 500
• 700
• 710 |

						<orgName> • one or more of <persName> or <orgName> followed by one <resp> Whenever possible, establish or use the form of the name from a national name authority file.		
		└	<meeting>	Optionally, record the name of a meeting or conference when this name is not clear from information in other parts of the <fileDesc>. Whenever possible, establish or use the form of the name from a national name authority file.	• 111 • 711	• 111 • 711
	├	<editionStmt>		This element contains information about the edition of the TEI document produced, not the source document.	250	n/a
	├	<publicationStmt>		Use the child elements below (rather than <p>) for a prose description.	n/a	n/a
		├	<publisher>	The publisher is the party responsible for making the file (the TEI document, not the source document) public.	• 260 $b • 533 $c*	n/a
		├	<distributor>	The distributor is the party from whom copies of the file (the TEI document, not the source document) can be obtained. Often the same as <publisher>, in which case no <distributor> should be given.	260 $b ($b is repeatable)	n/a
		├	<authority>	Only used for a text (the TEI document, not the source document) that is not formally published, but is nevertheless made available for circulation, in which case the party who makes it available should be recorded here.	500	n/a
		├	<idno>	Any unique identifier for the TEI document as determined by the publisher of the TEI document. Use of this element is recommended. Optionally use a **@type** attribute to indicate the type of identifier.	• 028 5_ • 099	n/a
		├	<availability><p>	Provide a prose rights statement for the TEI document. Provide a standard license, such as one from Creative Commons, if possible. Provide information on all applicable rights: rights in the original work, rights in page images of the source document, and rights in the encoded text.	540	n/a
		└	date when="____"/	Refers to the date of the first publication of the TEI document. Use the **@when** attribute (see [http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.datable.w3c.html att.datable.w3c class]) to aid machine processing. This element has no content.	• 260 $c • 533 $d*	n/a
	├	<seriesStmt>		This element contains information about the electronic series being created. It has one recommended	n/a	n/a

						element (<title>) and other optional elements.		
|	|	└ title level="s" type="_"				Required for the title of the series. Whenever possible, establish or use the form of the name from a national name authority file for the electronic series being created. Use of the @**type** attribute is optional, but if it is used, it should follow [http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-title.html instructions for use of this element in the full TEI Guidelines].	• 4xx • 8xx • 533 $f*	n/a
|	└ <notesStmt>					Optional.	5xx	5xx
|	└ <sourceDesc>					Use one <sourceDesc> per source document. Metadata for the source document may be automatically generated from a MARC record.	n/a	n/a
|		└ <biblStruct>				Use <biblStruct> with child elements arranged in the order below for ease of display according to ISBD. (This element is used instead of <bibl> to enforce structure, but <biblFull> is not used because it requires more elements than are typically available in library metadata sources.)	n/a	n/a
|			├ <analytic>			Use this element to group together elements describing the object of encoding when it would not have a corresponding catalog record—for example, an article in a journal issue, a chapter in a book, or a poem in a collection. '''If the object of encoding would have a corresponding catalog record, omit this element and its children.'''	n/a	n/a
|			|	├ <author>		One or more <author> elements (one name per element) are used to encode the name for the personal author or corporate body responsible for the creation of the intellectual or artistic content of the object of encoding. Use <persName> or <orgName> when applicable. Whenever possible, establish or use the form of the name from a national name authority file.	n/a	n/a
|			|	└ title level="a" type="_"		At least one <title> element is required for the title of the object of encoding. Transcribe the title according to the national cataloging code. Use of the @**type** attribute is recommended. It should have one of the following values as suitable in local practice: • main • sub • alt • short • desc • translated • filing (used for a version of the title with initial articles removed, to be used for sorting titles	n/a	n/a

						alphabetically but not for display)		
│		├		\<monogr\>		Use this element to group together the elements describing the bibliographic item that has (or would have) a corresponding catalog record. The TEI definition of this element specifies that it is used even for works that might not otherwise be considered "monographs," so bibliographic data about a journal title would be included in this element.	n/a	n/a
│		│	├	\<author\>		One or more \<author\> elements (one name per element) are used to encode the name for the personal author or corporate body responsible for the creation of the intellectual or artistic content of the source document bibliographic item, even if this creator is not the main entry in the catalog record. Use \<persName\> or \<orgName\> when applicable. Whenever possible, establish or use the form of the name from a national name authority file.	**MARC record based on encoded text** • 534 $a = 1st author **MARC record based on source document** • 100 • 110 • 700 • 710	• 100 • 110 • 700 • 710
│		│	├	title level="_" type="_"		At least one \<title\> element is recommended for the title of the source document bibliographic item. Transcribe the title according to the national cataloging code. Use of the **@level** attribute is optional. If used, it should be used [http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-title.html as in the main TEI Guidelines]. Use of the **@type** attribute is recommended. It should have one of the following values as suitable in local practice: • marc245a (used for the title proper and alternative title according to the national cataloging code) • filing (used for a version of the title with initial articles removed, to be used for sorting titles alphabetically but not for display) • marc245b (used for the the remainder of the title information — parallel titles, titles subsequent to the first, and other title information — according to the national cataloging code) • marc245c (used for the statement of responsibility according to the national cataloging code) • uniform (used for a uniform title according to the national cataloging code)	**MARC record based on encoded text** • 534 $t **MARC record based on source document** • 130 • 240 • 245 $a,$b • 246	• 130 • 240 • 245 $a,$b • 246
│		│	│	├	\<respStmt\>	Statement of responsibility on the source document bibliographic item,	245 $c	245 $c

									according to the national cataloging code. Record one responsibility or party per <respStmt>. Each <respStmt> should contain either: • one <resp> followed by one or more of <persName> or <orgName> • one or more of <persName> or <orgName> followed by one <resp> Whenever possible, establish or use the form of the name from a national name authority file. If generating the <sourceDesc> from a MARC record, it will be difficult to split the content of the 245c field into <resp> and <persName> (or <orgName>) elements, so it is recommended to use title type="marc245c" instead of this element.			
				├	<meeting>		Optionally, record the name of a meeting or conference when this name is not clear from information in other parts of the <sourceDesc>. Whenever possible, establish or use the form of the name from a national name authority file.	• 111 • 711	• 111 • 711
				├	<edition>		Edition statement (if present) according to the national cataloging code.	**MARC record based on encoded text** / **MARC record based on source document** • 534 $b / • 250	• 250
				├	<imprint>			n/a	n/a
					├	<pubPlace>	Place of publication from the source document bibliographic item according to the national cataloging code. Optionally remove ISBD punctuation for separating areas of the bibliographic description (such as a colon) when deriving from a MARC record. However, leave brackets that indicate supplied information or an abbreviation like "S.l." (for no place of publication).	**MARC record based on encoded text** / **MARC record based on source document** • 534 $c / • 260 $a	• 260 $a
					├	<publisher>	Name of publisher, distributor, etc. from the source document bibliographic item according to the national cataloging code. Optionally remove ISBD punctuation for separating areas of the bibliographic description (such as a comma) when deriving from a MARC record. However, leave brackets that indicate supplied information or an abbreviation like "s.n." (for no publisher).	**MARC record based on encoded text** / **MARC record based on source document** • 534 $c / • 260 $b	• 260 $b
					└	date when="____" *or* date	Date of publication, distribution, etc. from the source document bibliographic item. The content of the element is the statement of this data according to the national cataloging	**MARC record based on encoded text** / **MARC record based on source document**	260 $c

				`notBefore="____"` `notAfter="____"` *or* `date from="____"` *or* `date to="____"` *or* `date from="___"` `to="____"`	code. Since the content of the element according to the national cataloging code is not easily processed by machine, when possible include the following attribute(s) with valid values: "either" **@when**, "or" both **@notBefore** and **@notAfter**, "or" one or both of **@from** and **@to**. National cataloging codes may distinguish between a possible range of dates for publication (such as "186-" for something certainly published during the 1860s) and an uncertain date of publication (such as "1864?" or "186-?" for a date or range of dates assumed by the cataloger). In the case of uncertainty, use `cert="low"`. If the date is unknown (for example, recorded according to the national cataloging code as "[n.d.]", use `cert="unknown"`.	• 534 $c (content of element) • Dates fixed fields (value of attribute(s))	• 260 $c (content of element) • Dates fixed fields (value of attribute(s))	
|	||	|	└	`<extent>`	Use of this element to describe the extent of the source document bibliographic item is recommended. If the data is generated by hand, it should include a comprehensible statement of the size of the item, such as the number of pages or leaves. If generated from a catalog record, there should be two `<extent>` elements: one for the extent of the item (e.g., number of pages) and other physical details, and a second one for the dimension(s). Both should be recorded according to a national cataloging code.	**MARC record based on encoded text** • 534 $e	**MARC record based on source document** • 300	• 300
|	||	|	├	`<series>`	Information about the series to which the source document bibliographic item belongs, given according to the national cataloging code. If generating this data from a catalog record, it is likely that you will have only one child element: a title level="s". Use of the **@type** attribute *on the <title> element* is optional, but if it is used, it should follow instructions for use of this element in the full TEI Guidelines.	**MARC record based on encoded text** • 534 $f	**MARC record based on source document** • 4xx • 8xx	• 4xx • 8xx
|	||	|	├	`<note>`	Optionally, use for notes about the source document bibliographic item, given according to a national cataloging code.	**MARC record based on encoded text** • 534 $n	**MARC record based on source document** • 5xx	5xx
|	||	|	├	`<idno>`	Optionally use one or more `<idno>` elements to give identifiers for the source document, text, or work of the bibliographic item, whether assigned by the holding library (such as a call number), the publisher of the original document (such as an ISBN), or a standard bibliography (such as an identifier from the *Short Title Catalogue* or *Books in Maori*). Use the following values for the **@type** attribute if applicable, and create other	**MARC record based on encoded text** • 534 $z for ISBN	**MARC record based on source document** (possibly n/a) • 500 • 776 $w	• 015 • 016 • 020 • 024 • 025 • 027 • 028 • 029 • 035 • 050-099

							values if appropriate: • LC_call_number • isbn-13 • isbn-10		
|			└		\<relatedItem>		Use this element and its children to reference a related work, if applicable.	n/a	n/a
|				└	\<bibl>			n/a	n/a
|					├	\<author>	Optionally use one or more \<author> elements (one name per element) to encode the name for the personal author or corporate body responsible for the creation of the intellectual or artistic content of the related work. Use \<persName> or \<orgName> when applicable. Whenever possible, establish or use the form of the name from a national name authority file.	n/a	n/a
|					└	title type="_"	At least one \<title> element is recommended for the title of the related work. Transcribe the title according to the national cataloging code. Use of the **@level** attribute is recommended. If used, it should be used as in the main TEI Guidelines. Use of the **@type** attribute is optional. It should have one of the following values as suitable in local practice: • main • sub • alt • short • desc • translated • filing (used for a version of the title with initial articles removed, to be used for sorting titles alphabetically but not for display)	740	740
├	\<encodingDesc>							n/a	n/a
|	├	\<projectDesc>\<p>					Enter a description of the purpose for which the electronic file was encoded.	500	n/a
|	├	editorialDecl n="_"					Use of the **@n** attribute is recommended to record the encoding level: 1 for Level 1, 2 for Level 2, etc. Include one or more \<p> elements as children with information on: • editorial decisions made during encoding • notes about omissions of material found in the original work • the format of the data in the header: Does the data in the \<sourceDesc> follow AACR rules? How about in the \<fileDesc>? Is ISBD punctuation included? • automated processes used to generate the markup or content • external files or databases	• 500 for content of p element • 856 $z, which includes boilerplate text depending on encoding level and how the TEI document is presented to the user (as page images, text, or both)	n/a

				(such as those containing authority data) referenced in the TEI document • whether line breaks, column breaks, and/or page breaks are encoded • whether hyphens and quotation marks have been retained as character data or removed and indicated by the presence of an element such as <lb>, <cb>, <pb>, <quote>, or <floatingText> • whether types of hyphens have been distinguished (applies to Level 3 only)		
│	├		<tagsDecl>		n/a	n/a
│	│	├	rendition xml:id="_" scheme="css"	Include one or more <rendition> elements for each unique value of a **@rendition** attribute (not **@rend** attribute) used in the body of the TEI document. The **@xml:id** attribute is required in order to provide an identifier to which **@rendition** attributes in the body refer.	n/a	n/a
│	│	└	namespace name="http://www.tei-c.org/ns/1.0"<tagUsage>	<tagUsage> should be one of the following: • `<tagUsage gi="div1">Numbered divs used.</tagUsage>` • `<tagUsage gi="div">Unnumbered divs used.</tagUsage>`	n/a	n/a
│	└		<classDecl>taxonomy xml:id="____"<bibl>	Use to document classification schemes and controlled vocabularies referenced by a **@scheme** attribute elsewhere in the header or body of the TEI document. For example: • `<taxonomy xml:id="LCC"> <bibl>Library of Congress Classification</bibl> </taxonomy>` • `<taxonomy xml:id="LCSH"> <bibl>Library of Congress Subject Headings</bibl> </taxonomy>` • `<taxonomy xml:id="AAT"> <bibl>Art & Architecture Theasaurus</bibl> </taxonomy>` The **@xml:id** attribute is required in order to provide an identifier to which **@scheme** attributes in elsewhere in the header refer.	050-099 for call number classification schemes 6xx 2nd indicator or 6xx $2 when 2nd indicator = 7 for subject classification schemes	050-099 for call number classification schemes 6xx 2nd indicator or 6xx $2 when 2nd indicator = 7 for subject classification schemes
├			<profileDesc>		n/a	n/a
│	├		<langUsage>	Optionally use this element and child	008/35-37	n/a

				<language> elements to list languages used in the text. This supplements the **@xml:lang** attribute on the <text> (which is outside the header) in cases where more than one language is used in the text. It is not expected that the <langUsage> element will contain any description of language usage.		
		└	language ident="___"	Use one or more <language> elements to indicate language(s) used in the source document. Use of the **@ident** attribute is required as in the full TEI guidelines. Since the value of this attribute is usually sufficient to indicate the language, the <language> element should normally have no content. In the unusual case where **@ident** is insufficient, provide additional information about the language as content of the element.	• 041 • 546	• 041 • 546
	└	<textClass>			n/a	n/a
		├	classCode scheme="___"	True classification numbers as opposed to call numbers may be entered here. The value of the scheme attribute corresponds to a classification scheme defined previously in <classDecl>. Example: `scheme="#LCC"`	050-099	050-099
		└	keywords scheme="____"	Repeat this element as many times as there are keyword schemes. The value of the **@scheme** attribute is a URI for a controlled or uncontrolled vocabulary. The URI may be absolute to a version online or to one defined previously in <classDecl>. Example: `scheme="#LCSH"`	6xx 2nd indicator or 6xx $2 when 2nd indicator = 7	6xx 2nd indicator or 6xx $2 when 2nd indicator = 7
			└	<term>	Use for terms from controlled or uncontrolled vocabularies as defined according to the containing <keywords> element.	6xx	6xx
└	<revisionDesc>			n/a	n/a
	└	change when=""YYYY-MM-DD"" who=""URI""	Create a <change> element to record each significant change to the TEI document, in reverse chronological order (i.e., most recent first). A prose description of the change is recorded as the content of each <change> element. This prose may contain lists for organization, and phrase-level markup (like <gi>, <ptr>, or <date>), but not paragraphs. The date of the change should be recorded using the **@when** attribute (see att.datable.w3c class). The person who is responsible for making the change should be indicated by the **@who** attribute of <change>. Its value is a URI that points to a <respStmt> or <person> that encodes information about the responsible party. Note that this reference is a URI reference and not an ID/IDREF reference, and thus is not checked by validation software. Small projects sometimes take	n/a	n/a

| | | | advantage of this by putting information into the URI itself, and not having a <respStmt> or <person> element. For example, the document might simply give who="#Jane_Smith", relying on human readers to understand this reference. | | |

* Use only if TEI header metadata is based on the source document, not the encoded text.

4.1.7. Sample TEI Header

```
<teiHeader xml:lang="en">
 <fileDesc>
  <titleStmt>
   <title type="main">Lincoln and Seward.</title>
   <author>
    <persName>Welles, Gideon, 1802-1878.</persName>
   </author>
  </titleStmt>
  <publicationStmt>
   <publisher>University of Michigan, Digital Library Initiatives</publisher>
   <availability>
    <p>These pages may be freely searched and displayed. Permission must be received for
       subsequent distribution in print or electronically. Please go to
       http://www.umdl.umich.edu/ for more information.</p>
   </availability>
   <date when="1996"/>
  </publicationStmt>
  <seriesStmt>
   <title level="s" type="main">Making of America</title>
  </seriesStmt>
  <sourceDesc>
   <biblStruct>
    <monogr>
     <author>
      <persName>Welles, Gideon, 1802-1878.</persName>
     </author>
     <title level="m" type="marc245a">Lincoln and Seward.</title>
     <title level="m" type="marc245b">Remarks upon the memorial address of Chas. Francis
        Adams, on the late William H. Seward, with incidents and comments illustrative of the
        measures and policy of the administration of Abraham Lincoln. And views as to the
        relative positions of the late President and secretary of state.</title>
     <title type="marc245c">By Gideon Welles</title>
     <imprint>
      <pubPlace>New York</pubPlace>
      <publisher>Sheldon &amp; company</publisher>
      <date when="1874">1874</date>
     </imprint>
     <extent>viii, [7]-215 p</extent>
     <extent>20 cm.</extent>
    </monogr>
    <note>First published in condensed form in the Galaxy, v. 16, 1873, p. [518]-530,
       [687]-700, [793]-804.</note>
    <idno type="isbn-10">1-4255-1817-6</idno>
    <idno type="LC_call_number">E456 .W44</idno>
   </biblStruct>
  </sourceDesc>
 </fileDesc>
 <encodingDesc>
  <projectDesc>
   <p>XML created for the Making of America collection.</p>
  </projectDesc>
  <editorialDecl n="1">
   <p>Data in the <gi>sourceDesc</gi> of the header comes from a pre-AACR2 record. Other data
      follows AACR2 when applicable.</p>
   <p>
    <gi>sourceDesc</gi> created by exporting from catalog on 2008-06-15.</p>
   <p>This electronic text file was created by optical character recognition (OCR). No
      corrections have been made to the OCR-ed text and no editing has been done to the content
      of the original document. Encoding has been done using the recommendations for Level 1 of
      the TEI in Libraries Guidelines.</p>
   <p>Line breaks and column breaks have not been encoded, but page breaks have.</p>
   <p>All hyphens and quotation marks have been retained.</p>
  </editorialDecl>
  <tagsDecl>
   <namespace name="http://www.tei-c.org/ns/1.0">
    <tagUsage gi="div">Unnumbered divs used.</tagUsage>
   </namespace>
```

```
    </tagsDecl>
   <classDecl>
    <taxonomy xml:id="LCC">
     <bibl>Library of Congress Classification</bibl>
    </taxonomy>
    <taxonomy xml:id="LCSH">
     <bibl>Library of Congress Subject Headings</bibl>
    </taxonomy>
   </classDecl>
  </encodingDesc>
  <profileDesc>
   <langUsage>
    <language ident="en"/>
   </langUsage>
   <textClass>
    <classCode scheme="#LCC">E456</classCode>
    <keywords scheme="#LCSH">
     <list>
       <item>Lincoln, Abraham, 1809-1865.</item>
       <item>Seward, William Henry, 1801-1872.</item>
       <item>Adams, Charles Francis, 1807-1886. Address of Charles Francis Adams ... on the life
             ... of William H. Seward.</item>
     </list>
    </keywords>
   </textClass>
  </profileDesc>
  <revisionDesc>
   <change who="#CKP" when="2005-05-25">Header generated from export of MARC record</change>
  </revisionDesc>
 </teiHeader>
```

4.1.8. Specification

The <extent> element should not be used as a direct child of <fileDesc>, but rather only as a descendant of a bibliographic citation (e.g., of the source). When inside an <imprint>, a <date> element should have a machine-readable version of the date specified either on the @when attribute, or, in the case of not knowing the precise date, on the @notBefore and @notAfter attributes. The use of the = attribute (found here on the <> element) is not recommended by the Best Practices for TEI in Libraries number Used to indicate which encoding level described by *TEI in Libraries: Guidelines for Best Practices* is in use. 12345

```
<editorialDecl n="1">
 <p>Metadata in the TEI header comes from an AACR2-conformant
   record, translated to TEI via the <name type="software">Thutmose I</name> program.</p>
 <p>Content originally generated by <name type="software">c-n-rite</name> OCR software, then the
   needed TEI encoding put in place with <name type="software">cnr2tei.xslt</name>.</p>
 <p>All hyphens in source document encoded as U+2010.</p>
</editorialDecl>
```

The TEI provides a set of useful special-purpose elements that can be used inside of <editorialDecl> instead of paragraphs: <correction>, <hyphenation>, <interpretation>, <normalization>, <quotation>, <segmentation>, and <stdVals>. But since at the time these *Best Practices* were developed one could not use any of these special-purpose elements *and* a paragraph, and there are some editorial practices likely to be of interest that are not covered by these elements, our current requirement is to use only paragraphs. The TEI has since fixed this bug (by allowing a mix of paragraphs and the special-purpose elements in any order), this recommendation will likely change accordingly in the near future.

It is required that a paragraph explaining hyphenation practices, with particular wording as above, be present.

U.S. Library of Congress call number intnerational serial book number, 13-digit intnerational serial book number, 10-digit http://www.tei-c.org/ns/1.0 Use of specialized child elements of the publication statement (rather than paragraphs) is recommended whenever possible Dates inside the publication statement must have @when (and should not have content) Dates inside the publication statement should not have content (and must have @when)

For normally published items, the specialized children elements (e.g. <publisher>) should be used. Paragraphs are permitted as an alternative for unusual cases like unpublished works.

Cascading Stylesheet Language Unnumbered divs used.Numbered divs used. element name the name (generic identifier) of the element indicated by the tag. A single TEI in Libraries document should not mix numbered and unnumbered divisions. The use of divisions (i.e., whether numbered or unnumbered divisions are used) must be documented in a <tagUsage> element (inside a <namespace> element ... A document that uses un-numbered divisions should specify so by using <tagUsage gi="div">Unnumbered divs used.</tagUsage>; a document that uses numbered divisions should specify so by using <tagUsage gi="div1">Numbered divs used.</tagUsage>. divisions are un-numbered The document uses <div>, not <div1>, <div2>, etc. The content of this <tagUsage> should be 'Unnumbered divs used.'. divisions are numbered The document uses <div1> (and perhaps <div2>, etc.), not <div> elements. The content of this <tagUsage> should be 'Numbered divs used.'. indicates the bibliographic level for a title, that is, whether it identifies an article, book, journal, series, or unpublished material. The @level attribute should not be specified on a <title> within the <titleStmt>. When a child of <seriesStmt>, the level of a title must be specified as 's'. When inside the series-level portion of a structured bibliographic citation, the level of a title must be specified as 's'. When a child of <analytic>, the level of a title must be specified as 'a'. When a child of <monogr>, the level of a title must not be specified as 'a'. analytic analytic title (article, poem, or other item published as part of a larger item) monographic monographic title (book, collection, or other item published as a distinct item, including single volumes of multi-volume works) journal journal title series series title unpublished title of unpublished material (including theses and dissertations unless published by a commercial press)

The level of a title is sometimes implied by its context: for example, a title appearing directly within an <analytic> element is *ipso facto* of level a, and one appearing within a <series> element of level s. For this reason, the **@level** attribute is not required in contexts where its value can be unambiguously

inferred. Where it is supplied in such contexts, its value should not contradict the value implied by its parent element.

Inside the title statment the @type of a <title> should be specified Inside a structured bibliographic citation of a mongraphic-level item, @type of a <title> should be specified Inside the analytic portion of a structured bibliographic citation, @type of a <title> should be specified Inside a bibliographic citation of a related item, the @type of a <title> should be specified main title subordinate subtitle, title of part alternate alternate title, often in another language, by which the work is also known abbreviated form of title descriptive a translation of a title used for the title proper and alternative title according to the national cataloging code used for the the remainder of the title information — parallel titles, titles subsequent to the first, and other title information — according to the national cataloging code) used for the statement of responsibility according to the national cataloging code used for a uniform title according to the national cataloging code

4.2. Encoding Levels

4.2.1. Caveats About Examples

In the examples given in the description of each encoding level below, XML comments are illustrative, and are not meant to be included in encoded documents. Here is an example of such a comment:

```
<!-- uncorrected OCR for first page image begins here -->
```

Note that for technical reasons the namespace is not shown in these examples, but it should always be supplied on the root <TEI> element, e.g.:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
```

4.2.2. Level 1: Fully Automated Conversion and Encoding

4.2.2.1. Reference

- Chapter 3, Elements Available in All TEI Documents

Note that this is not a 'TEI conformant' customization, because it does not follow the TEI abstract model. However, this is a 'syntactically conformant' customization, in that documents that are valid against this scheme will also be valid against the TEI_all schema.

4.2.2.2. Purpose

To create electronic text with the primary purpose of keyword searching and linking to page images. The primary advantage in using the TEI at this very strictly limited level of encoding is that a TEI header is attached to the text file.

4.2.2.3. Rationale

The text is subordinate to the page image, and is not intended to stand alone as an electronic text (without page images). Level 1 texts are not intended to be adequate for textual analysis; they are more likely to be suited to the goals of a preservation unit or mass digitization initiative. Though their encoding is minimal, Level 1 texts are fully valid XML texts. In addition to taking advantage of the TEI header, these texts, while lightly encoded, can be easily combined with more richly encoded texts (that also follow these guidelines) for searching. Further encoding based on document structures or content analysis can be added to a Level 1 text at any time.

Level 1 is most suitable for projects with the following characteristics

- a large volume of material is to be made available online quickly
- a digital image of each page is desired
- no manual intervention will be performed in the text creation process
- the material is of interest to a large community of users who wish to read texts that allow keyword searching
- sophisticated search and display capabilities based on the structure of the text are not necessary
- extensibility is desired; that is, one desires to keep open the option for a higher level of encoding to be added at a later date

4.2.2.4. Workflow

Texts at Level 1 can be created and encoded by fully automated means. Page images are scanned and processed using OCR, but the text is left uncorrected ("dirty OCR"). Page images are tagged using software that assigns page-level metadata (page number and possibly tags for page features) to each page image for display in the user interface in a list of pages. Encoding is performed automatically: markup with page-level metadata is inserted at selected points into the dirty OCR text, generating a valid XML document. This encoding is both minimal and reliable, and does not typically require extensive review of each page of each text.

4.2.2.5. Element Recommendations for Level 1

<div> or <div1>	There should be only one child of <body>: a single <div> (or <div1>).
<ab>	There should be only one child of the <div> (or <div1>): a single <ab> wrapping all of the OCR text. If the text is ever "upgraded" to Level 3 or higher, the <ab> element will be replaced by structural elements like <p> and <table>.
<pb> or <facsimile>	See the explanation above for how to link between the encoded text and images of source documents. If using <pb>, it is recommended to put the element within an <ab> element.

4.2.2.6. Level 1 Example: Alger Hiss document

```
<TEI xml:id="someid">
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
 <text xml:lang="en">
  <body>
   <div1>
    <ab>
     <pb n="113" facs="00000001.tif"/>
        <!-- uncorrected OCR for first page image begins here -->
        POINT VIII.
        BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
        CONVICTION SHOULD BE VACATED; ALTERNATIVELY,
        DISCOVERY AND A HEARING SHOULD BE ORDERED.
        The nature and extent of surveillance of Hiss, his
        family and associates was not known at the time of trial by
        the defense. Even now, with the release of some of the govern-
        ment documents concerning FBI investigative techniques regarding
        Hiss, the full extent of surveillance -- wiretapping, mail open-
        ings, mail covers, physical surveillance, and other intrusive
        techniques -- is still not 'clear. Nevertheless, it is apparent
        that information gathered through the exploitation of unlawful
        wiretaps and other illegal surveillance was used at trial and
        consequently the conviction must be reversed. Alternatively,
        further discovery and a hearing is essential to a fair deter-
        mination regarding these issues.
        FBI surveillance of Hiss began in earnest in 1941 with
        the institution of a mail cover on his incoming correspondence
        at his home in connection with an FBI investigation of possible
        Hatch Act violations. CN Ex. 98A. Another mail cover was placed


        -113 -


        <!-- uncorrected OCR for first page image ends here -->
     <pb n="114" facs="00000002.tif"/>
        <!-- uncorrected OCR for second page image begins here -->
        on the Hiss mail in 1945, and at the same time the FBI obtained
        toll call records from the Hiss residence Telephone for the
        years 1943 and 1944 as well. CN Ex. 99. In September, 1945,
        the FBI intercepted telegrams to Hiss as well. CN Ex. 100.
        In late November, 1945, FBI surveillance of the Hiss
        residence in Washington, D.C., escalated. For the third time,
        a mail cover was instituted beginning on November 28, 1945,
        which was continued at least until 1946. CN Ex. 101 at p. 70;
        CN Ex. 102. Continuous physical surveillance of Hiss was begun
        as well. CN Ex. 101 at p. 72. Although this twenty-four-hour
        surveillance was discontinued on December 14, 1945, physical
        surveillance was conducted frequently at various times until
        September, 1947. CN Ex. 102; CN Ex. 103.
        The most intrusive invasion of petitioner's rights
        68/ Also before 1947, a letter from Priscilla Hiss addressed
        to her son, Timothy Hobson, was intercepted and its contents
        read. CN Ex. 100A at p. 167. In approximately March, 1947,
        a letter from a Michael Greenberg addressed to petitioner re-
        garding an application for employment with the United Nations
        was also intercepted, in a manner not revealed by the docu-
        ments. CN Ex. 100B


        -114 -


        <!-- uncorrected OCR for second page image ends here -->
     <pb n="115" facs="00000003.tif"/>
```

```
        <!-- uncorrected OCR for third page image begins here -->
        occurred from December 13, 1945 until the Hisses moved from
        Washington, D.C. to New York City on September 13, 1947. A
        "technical surveillance," -- a wiretap -- was placed on the Hiss
        telephone at their residence on P Street-in Washington, D.C.
        The logs of this surveillance constitute twenty-nine volumes
        of FBI serials and are roughly 2,500 pages in length, in which
        an enormous amount of information concerning the Hisses' per-
        sonal lives, relationships with friends and associates, and
        habits is recorded.
        The wiretap was installed following FBI Director Hoover's
        application to the Attorney General for authorization, although
        no written authorization appears in the documents released to
        Hiss. The purpose of the application was to gather information
        regarding Hiss' alleged contacts with Soviet espionage agents and
        communists in government service, general allegations which had
        been made by Elizabeth Bentley and Chambers.
        As one would expect, the interception of every telephone
        h9/ Hoover's initial request was answered by a note reques-
        ting information on Hiss. CN Ex. 104. Additional information
        was furnished by letter dated November 30, 1945. CN Ex. 105.


        -115 -


        <!-- uncorrected OCR for third page image ends here -->
      </ab>
    </div1>
   </body>
  </text>
 </TEI>
```

4.2.2.7. Specification

contains a single TEI-in-Libraries level 1 document, comprising a TEI header and a text, either in isolation or as part of a <teiCorpus> element.

For technical reasons, the TEI namespace is not displayed in examples. However, a TEI namespace declaration is required. It is typically given once on the TEI root element, e.g. TEI xmlns="http://www.tei-c.org/ns/1.0".

```
<TEI>
 <teiHeader>
  <fileDesc>
   <titleStmt>
    <title>A Short Level 1 Document</title>
   </titleStmt>
   <publicationStmt>
    <p>Only published as an example.</p>
   </publicationStmt>
   <sourceDesc>
    <p>Since this is an example, it doesn't really have a source</p>
   </sourceDesc>
  </fileDesc>
 </teiHeader>
 <text>
  <body>
   <div>
    <ab>This is about the shortest TEI document imaginable.</ab>
   </div>
  </body>
 </text>
</TEI>
```

This element is required. The TEI namespace should be specified on this element, e.g. TEI xmlns="http://www.tei-c.org/ns/1.0".
contains the entire content of the document; or, when used within the <teiHeader>, contains any arbitrary component-level unit of text, acting as an anonymous container for phrase- or inter-level elements analogous to, but without the semantic baggage of, a paragraph.

```
<publicationStmt>
 <availability>
  <ab>Copyleft 2009 Syd Bauman</ab>
 </availability>
</publicationStmt>
```

```
<body>
```

```
  <div>
   <ab> In the beginning God created the heaven and the earth.
      And the earth was without form, and void; and darkness was upon
      the face of the deep. And the spirit of God moved upon the face
      of the waters.
      And God said, Let there be light: and there was light.
   </ab>
  </div>
 </body>
```

At level 1, the entire document is encoded as a single <ab> — one and only one <ab> must be present within <text>. Further, <ab> *may* be used in a variety of places within the <teiHeader>.

in a bibliographic reference, contains the name (typically encoded as <name>, <persName>, or <orgName>) of the author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.

```
<author>
 <orgName>British Broadcasting Corporation</orgName>
</author>
```

```
<author>
 <persName ref="#mdalmau.cny">Michelle Dalmau</persName>
</author>
```

```
<author>
 <name>Matthew Gibson</name>
</author>
```

```
<author>anonymous</author>
```

```
<author>unknown</author>
```

contains the entirety of a single unitary text, including any front or back matter.

```
<body>
 <div>
  <ab> Nu scylun hergan hefaenricaes uard
      metudæs maecti end his modgidanc
      uerc uuldurfadur sue he uundra gihuaes
      eci dryctin or astelidæ
      he aerist scop aelda barnum
      heben til hrofe haleg scepen.
      tha middungeard moncynnæs uard
      eci dryctin æfter tiadæ
      firum foldu frea allmectig
      primo cantauit Cædmon istud carmen.
  </ab>
 </div>
</body>
```

At level 1, the content of <body> must be a single <div> or <div1> element.

At level 1, the content of <div> must be a single <ab> element.

section

At level 1, the content of <div1> must be a single <ab> element.

contains the name (typically encoded as <name>, <persName>, or <orgName>) of an individual, institution, or organization acting as editor.

```
<editor>
 <persName ref="names.xml#khawkins.tvt">Kevin Hawkins</persName>
</editor>
```

4.2.3. Level 2: Minimal Encoding

4.2.3.1. Reference

- Chapter 3, Elements Available in All TEI Documents
- Chapter 4, Default Text Structure

Note that this is a 'syntactically conformant' customization, in that documents that are valid against this scheme will also be valid against the TEI_all schema. However, it is unkown whether or not it is truly 'TEI conformant', as the *TEI Guidelines* do not make clear whether or not encoding of individual paragraphs is mandatory.

4.2.3.2. Purpose

To create electronic text for full-text searching, linking to page images, and identifying simple structural hierarchy to improve navigation. (For example, you can create a table of contents from such encoding.)

4.2.3.3. Rationale

The text is mainly subordinate to the page image, though navigational markers (textual divisions, headings) are captured. However, the text could stand alone as electronic text (without page images) if the accuracy of its contents is suitable to its intended use and it is not necessary to display low-level typographic or structural information. Use cases for Level 2 require a set of elements more granular than those of Level 1, including bibliographic or structural information below the monographic or volume level. One of the motivations for using Level 2 is to avoid expensive analysis of textual elements and/or the expense of accurate text conversion, e.g., double-keying or detailed proofreading of automatic OCR.

For the most part, Level 2 texts are *not* intended to be displayed separately from their page images. Level 2 encoding of sections and headings provides greater navigational possibilities than Level 1 encoding, and enables searching to be restricted within particular textual divisions (for example, searching for two phrases within the same chapter).

Level 2 is most suitable for projects in which

- a large volume of material is to be made available online quickly
- a digital image of each page is desired
- the material is of interest to a large community of users who wish to read texts that allow keyword searching
- rudimentary search and display capabilities based on the large structures of the text are desired
- each text is checked to ensure that textual divisions and headers are properly identified
- extensibility is desired; that is, one desires to keep open the option for a higher level of encoding to be added at a later date

4.2.3.4. Workflow

Level 2 generally can be created and encoded by automated means. Pagination is identified as in Level 1, and metadata for the textual divisions is created, likely based on the page images. The textual division metadata might contain the page number on which the division begins and a transcription of that division's heading. This metadata is inserted into the raw OCR at the appropriate points, forming a valid XML document. Level 2 texts do not require any special knowledge or manual intervention below the section level.

4.2.3.5. Element Recommendations for Level 2

Use all elements specified in Level 1 *plus the following*:

<front>, <back>	Optional. Contains one or more <div> or <div1>.
<body>	Contains one or more <div> or <div1>.
<div1> or <div>	Unlike in Level 1, in Level 2 one <div> or <div1> is used per section of the text identified with division-level metadata. If no **@type** attribute is specified, a **@type** value of section should be presumed.
<head>	Recommended if headings are present. As in the TEI, this element must be the first child of a <div> or <div1>.

4.2.3.6. Level 2 Examples

Note that for technical reasons the namespace is not shown in these examples, but it should always be supplied on the root <TEI> element, e.g.: TEI xmlns="http://www.tei-c.org/ns/1.0".

4.2.3.6.1. Level 2 Basic Structure

```
<TEI>
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
```

```
 <text xml:lang="en">
  <front> <!-- entire <front> is optional, but it might have, for example: -->
   <div type="titlePage">
     <pb facs="[URI of title page image]"/>
     <ab>[ entire title page here ]</ab>
   </div>
   <div type="TOC">
     <pb n="ii" facs="[URI of table of contents]"/>
     <head>[ heading of table of contents ]</head>
     <ab>[ entire table of contents here ]</ab>
   </div>
   <div type="preface">
     <head>[ heading of preface ]</head>
     <ab>[ entire preface, with interspersed <pb/> elements pointing
         to page images as needed, here ]</ab>
   </div>
  </front>
  <body>
   <div type="section">
    <pb n="1" facs="[URI of page 1 image]"/>
    <head>[ heading of section 1 ]</head>
    <ab>[ entire contents of section 1 here, with
         interspersed <gi>pb</gi> elements pointing to page
         images; in this example there are 26 more pages
         to section 1 ]</ab>
   </div>
   <div type="section">
    <pb n="27" facs="[URI of page 27 image]"/>
    <div type="subsection">
     <head>[ heading of section 2 subsection 1 ]</head>
     <ab>[ all the paragraphs of subsection one go here
          with page breaks inserted ]</ab>
    </div>
   </div>
  </body>
  <back> <!-- optional: organized like <front>, with 1 or more <div> or <div1>,
      each with a single <ab> -->
  </back>
 </text>
</TEI>
```

4.2.3.6.2. Level 2 Alger Hiss document

```
<TEI xml:id="someid">
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
 <text xml:lang="en">
  <body>
   <div1>
    <pb n="113" facs="00000001.tif"/>
       <!-- content of head element was transcribed from page image -->
   <head>POINT VIII. BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
        CONVICTION SHOULD BE VACATED; ALTERNATIVELY, DISCOVERY AND A HEARING
        SHOULD BE ORDERED.</head>
   <ab> <!-- uncorrected OCR for first page image begins here -->
        POINT VIII.
        BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
        CONVICTION SHOULD BE VACATED; ALTERNATIVELY,
        DISCOVERY AND A HEARING SHOULD BE ORDERED.
        The nature and extent of surveillance of Hiss, his
        family and associates was not known at the time of trial by
        the defense. Even now, with the release of some of the govern-
        ment documents concerning FBI investigative techniques regarding
        Hiss, the full extent of surveillance -- wiretapping, mail open-
        ings, mail covers, physical surveillance, and other intrusive
        techniques -- is still not 'clear. Nevertheless, it is apparent
        that information gathered through the exploitation of unlawful
        wiretaps and other illegal surveillance was used at trial and
        consequently the conviction must be reversed. Alternatively,
        further discovery and a hearing is essential to a fair deter-
        mination regarding these issues.
        FBI surveillance of Hiss began in earnest in 1941 with
        the institution of a mail cover on his incoming correspondence
        at his home in connection with an FBI investigation of possible
        Hatch Act violations. CN Ex. 98A. Another mail cover was placed
```

```
                     -113 -


                     <!-- uncorrected OCR for first page image ends here -->
          <pb n="114" facs="00000002.tif"/>
                     <!-- uncorrected OCR for second page image begins here -->
                     on the Hiss mail in 1945, and at the same time the FBI obtained
                     toll call records from the Hiss residence Telephone for the
                     years 1943 and 1944 as well. CN Ex. 99. In September, 1945,
                     the FBI intercepted telegrams to Hiss as well. CN Ex. 100.
                     In late November, 1945, FBI surveillance of the Hiss
                     residence in Washington, D.C., escalated. For the third time,
                     a mail cover was instituted beginning on November 28, 1945,
                     which was continued at least until 1946. CN Ex. 101 at p. 70;
                     CN Ex. 102. Continuous physical surveillance of Hiss was begun
                     as well. CN Ex. 101 at p. 72. Although this twenty-four-hour
                     surveillance was discontinued on December 14, 1945, physical
                     surveillance was conducted frequently at various times until
                     September, 1947. CN Ex. 102; CN Ex. 103.
                     The most intrusive invasion of petitioner's rights
                     68/ Also before 1947, a letter from Priscilla Hiss addressed
                     to her son, Timothy Hobson, was intercepted and its contents
                     read. CN Ex. 100A at p. 167. In approximately March, 1947,
                     a letter from a Michael Greenberg addressed to petitioner re-
                     garding an application for employment with the United Nations
                     was also intercepted, in a manner not revealed by the docu-
                     ments. CN Ex. 100B


                     -114 -


                     <!-- uncorrected OCR for second page image ends here -->
          <pb n="115" facs="00000003.tif"/>
                     <!-- uncorrected OCR for third page image begins here -->
                     occurred from December 13, 1945 until the Hisses moved from
                     Washington, D.C. to New York City on September 13, 1947. A
                     "technical surveillance," -- a wiretap -- was placed on the Hiss
                     telephone at their residence on P Street-in Washington, D.C.
                     The logs of this surveillance constitute twenty-nine volumes
                     of FBI serials and are roughly 2,500 pages in length, in which
                     an enormous amount of information concerning the Hisses' per-
                     sonal lives, relationships with friends and associates, and
                     habits is recorded.
                     The wiretap was installed following FBI Director Hoover's
                     application to the Attorney General for authorization, although
                     no written authorization appears in the documents released to
                     Hiss. The purpose of the application was to gather information
                     regarding Hiss' alleged contacts with Soviet espionage agents and
                     communists in government service, general allegations which had
                     been made by Elizabeth Bentley and Chambers.
                     As one would expect, the interception of every telephone
                     h9/ Hoover's initial request was answered by a note reques-
                     ting information on Hiss. CN Ex. 104. Additional information
                     was furnished by letter dated November 30, 1945. CN Ex. 105.


                     -115 -


                     <!-- uncorrected OCR for third page image ends here -->
          </ab>
        </div1>
      </body>
    </text>
  </TEI>
```

4.2.3.7. Specification

TEI document contains a single TEI-in-Libraries level 2 document, comprising a TEI header and a text, the latter represented as either a transcription (in <text>) or a transcription and page images (in <text> and <facsimile> respectively), either in isolation or as part of a <teiCorpus> element.

```
<TEI>
 <teiHeader>
  <fileDesc>
   <titleStmt>
    <title>A Short Level 2 Document</title>
```

```
    </titleStmt>
    <publicationStmt>
     <p>Only published as an example.</p>
    </publicationStmt>
    <sourceDesc>
     <p>Since this is an example, it doesn't really have a source</p>
    </sourceDesc>
   </fileDesc>
  </teiHeader>
  <text>
   <body>
    <div>
     <ab>This is about the shortest TEI document imaginable.</ab>
    </div>
   </body>
  </text>
 </TEI>
```

This element is required. The TEI namespace should be specified on this element, e.g. TEI xmlns="http://www.tei-c.org/ns/1.0".
anonymous block contains the entire content of a division of the document; or, when used within the <teiHeader>, contains any arbitrary component-
level unit of text, acting as an anonymous container for phrase or inter level elements analogous to, but without the semantic baggage of, a paragraph.

```
<publicationStmt>
 <availability>
  <ab>Copyleft 2009 Syd Bauman</ab>
 </availability>
</publicationStmt>
```

```
<body>
 <div>
  <head>Genesis</head>
  <ab> In the beginning God created the heaven and the earth.
     And the earth was without form, and void; and darkness was upon
     the face of the deep. And the spirit of God moved upon the face
     of the waters.
     And God said, Let there be light: and there was light.
  </ab>
 </div>
</body>
```

At level 2, entire sections of the document (be they parts, chapters, etc.) are each encoded as a single <ab> — each division (whether <div>, <div1>,
<div2>, etc.) should have one and only one <ab> child. Further, <ab> *may* be used in a variety of places within the <teiHeader>.
in a bibliographic reference, contains the name (typically encoded as <name>, <persName>, or <orgName>) of the author, personal or corporate, of a
work; for example in the same form as that provided by a recognized bibliographic name authority.

```
<author>
 <orgName>British Broadcasting Corporation</orgName>
</author>
```

```
<author>
 <persName ref="persons.xml#mdalmau.cny">Michelle Dalmau</persName>
</author>
```

```
<author>
 <name>Gibson, Matthew</name>
</author>
```

```
<author>anonymous</author>
```

```
<author>unknown</author>
```

Particularly where cataloguing is likely to be based on the content of the header, it is advisable to use a generally recognized name authority file to
supply the content for this element. The attributes **@key** or **@ref** may also be used to reference canonical information about the author(s) intended from
any appropriate authority, such as a library catalogue or online resource.
In the case of a broadcast, use an <orgName> inside this element for the name of the company or network responsible for making the broadcast.

Where an author is unknown or unspecified, this element may contain text such as *Unknown* or *Anonymous*.
[[undefined COBICOR]] [[undefined HD21]]

```
<back>
 <div1 type="appendix">
  <head>The Golden Dream or, the Ingenuous Confession</head>
  <ab>To shew the Depravity of human Nature </ab>
 </div1>
 <div1 type="epistle">
  <head>A letter from the Printer, which he desires may be inserted</head>
  <ab>Sir.
    I have done with your Copy, so you may return it to the Vatican, if you please</ab>
 </div1>
 <div1 type="advert">
  <head>The Books usually read by the Scholars of
    Mrs Two-Shoes are these and are sold at Mr
    Newbery's at the Bible and Sun in St Paul's
    Church-yard.</head>
  <ab> The Christmas Box, Price 1d.
    The History of Giles Gingerbread, 1d.
    A Curious Collection of Travels, selected from the Writers of all Nations,
    10 Vol, Pr. bound 1l.
  </ab>
 </div1>
 <div1 type="advert">
  <head>By the KING's Royal Patent,
    Are sold by J. NEWBERY, at the
    Bible and Sun in St. Paul's Church-Yard.</head>
  <ab> Dr. James's Powders for Fevers, the Small-Pox, Measles, Colds, &amp;c.
    2s. 6d
    Dr. Hooper's Female Pills, 1s.
  </ab>
 </div1>
</back>
```

```
<body>
 <head>LA FOREST
   NVPTIALE,</head>
 <ab>Où eſt repreſentee vne varieté bigarree,
   non mois eſmerveillable que plaiſan-
   te, de diuers mariages, ſelon qu'ils ſont
   obſerueZ &amp; pratiqueZ par pluſieurs
   peuples &amp; nations eſtranges. Auec
   la maniere de policier, regir, gouuer-
   ner &amp; adminiſtrer leur famille.</ab>
 <div type="chapter">
  <head>Les Romains. CHAPITRE I.</head>
  <ab>Encores que ie ne
    veuille me formali-
    ſer contre le droict
    Romain, neātmoins
    puis que le forma-
    litez, qui eſtoient an-
    ciennement gardees au nopces
    Romaines, ſont maintenant miſes
    hors d'vſages &amp; pratique, ie ne ſe-
    A
  <pb/>
    ray point de difficulté d'emprunter
    des anciens autheurs ce qui appar-
    <!-- ... -->
  </ab>
 </div>
</body>
```

At level 2, the content of <body> may contain only <head>, <ab>, <pb>, <note>, and either <div> or <div1> elements.
section

At level 2, <div> may contain only an optional <head> followed by a required <ab>; also <pb> and <note> elements may be interspersed anywhere as needed.
section

At level 2, <div1> may contain only an optional <head> followed by a required <ab>.
contains the name (typically encoded as <name>, <persName>, or <orgName>) of an individual, institution, or organization acting as editor.

```
<editor>
 <persName ref="names.xml#khawkins.tvt">Kevin Hawkins</persName>
```

```
  </editor>
```

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

```
<front>
 <div type="dedication">
  <ab>To our three selves</ab>
 </div>
 <div type="preface">
  <head>Author's Note</head>
  <ab>All the characters in this book are purely
     imaginary, and if the author has used names that
     may suggest a reference to living persons she has
     done so inadvertently. ...</ab>
 </div>
</front>
```

The <head> element is used for headings at all levels; software which treats (e.g.) chapter headings, section headings, and list titles differently must determine the proper processing of a <head> element based on its structural position.

A <head> occurring as the first element of a <div> or <div1> is the title of that chapter or section.

4.2.4. Level 3: Simple Analysis

4.2.4.1. Reference

- Chapter 4, Default Text Structure, P5 Guidelines
- Chapter 6, Verse, P5 Guidelines
- Chapter 14, Tables, Formulæ, and Graphics, P5 Guidelines
- Chapter 16, Linking, Segmentation, and Alignment, P5 Guidelines (for handling notes)

Note that this is intended to be a 'TEI conformant' customization, per P5 section 23.3.

4.2.4.2. Purpose

To create a stand-alone electronic text and identify hierarchy (logical structure) and typography without content analysis being of primary importance.

4.2.4.3. Rationale

Encoding at this level provides the foundation for upgrading to higher levels of encoding. Level 3 generally requires some human editing, but the features to be encoded are determined by the logical structure and appearance of the text and not specialized content analysis.

Level 3 texts identify front and back matter, textual divisions, and all paragraph breaks. Floating texts, or sub-texts like a poem or letter embedded in the greater text, are supported in this level. The finer granularity of encoding these features, as well as figures, notes, and all changes of typography, allows a range of options for display, delivery, and searching. For example, one has the option of identifying, and therefore specifying, the display characteristics of different typographic styles, and regularizing the display and placement of note text.

Level 3 texts can stand alone as text without page images, and therefore can be uploaded, downloaded, and delivered quickly, and require less storage space than digital collections with page images. However, the simple level of structural analysis and absence of specialized content analysis reflected in Level 3 encoding may make it desirable for some, depending on project priorities, to include page images in order to provide users with a fuller set of resources.

Level 3 is most suitable for projects with the following characteristics

- the material is of interest to a large community of users who wish to read texts that allow for keyword searching
- some sophistication of display, delivery, and searching based on structure of the text is desired
- each text will undergo quality control to ensure that encoding decisions have been made appropriately
- the users of the texts may have limited storage or display capabilities
- the creator of the texts has limited or no ability to provide content expertise to analyze, tag, or review texts
- extensibility is desired; that is, one desires to keep open the option for a higher level of encoding to be added at a later date

4.2.4.4. Workflow

Level 3 texts can be created by conversion from an electronic source such as an HTML file or word-processor document or from a print source, either through OCR or keyboarding. They can be generated trivially by converting from outsourced double-keyboarded texts conforming to TEI Tite, though some granularity of encoding will be lost in the translation.

4.2.4.5. Element Recommendations for Level 3

Use all elements specified in Levels 1 and 2 *except <ab>, plus the following*:

<front>, <back>	Recommended if present.
<div> or <div1>	At least one is recommended within each of <front>, <body>, and <back>; **@type** attribute is recommended.
<p>	Recommended for paragraph breaks in prose.
<lg> and <l>	Recommended for identifying groups of lines and lines, respectively.
<figure> and appropriate child elements	Recommended to refer to illustrative images and descriptive information about those images.
<floatingText>	Optionally used to indicate a floating text.
<note>	Recommended for notes.
<ptr> and <ref>	If a table of contents is encoded, recommended for linking to sections of the document. If notes are encoded at the point they occur in the text or at another point convenient when converting from a born-digital source document, recommended for encoding the point of reference.
<hi>	Recommended to indicate changes in typeface; **@rend** attribute is optional.
<list> and <item>	Optionally used to indicate ordered and unordered list structures.
<table>, <row>, and <cell>	Optionally used to indicate table structures.
<lb>	Optionally used to indicate line breaks.
<cb>	Optionally used to indicate column breaks.

4.2.4.6. General Level 3 Recommendations

4.2.4.6.1. Forme Work

Running heads, catch words, page numbers, signatures, and other artifacts derived from printing should not be included in Level 3, with the exception of page numbers, which are recorded using the **@n** attribute on <pb>. If upgrading a text from Level 1 or Level 2 that was generated using OCR, discard the forme work information.

4.2.4.6.2. Level 3 Figures

<figure> groups elements representing or containing graphic information such as an illustration or figure; in this context <figure> contains the following elements:

<head>
  for a caption label (e.g., 'Figure 1') and/or a literal transcription of a caption. Use when this feature is present in the source document.
<p>
  for a literal transcription of a caption (could be used in conjunction with the <head> tag if a caption label is present). Use when this feature is present in the source document.
<figDesc>
  for free text description of the image for use when documenting an image without displaying it. This is *mandatory* in order to create digital texts that will be accessible to the visually impaired.
<graphic>
  for pointing to the URI of the image itself using a **@url** attribute and containing other presentation instructions such as dimension at which the graphic should be displayed, etc. This is *mandatory* in order to point to the corresponding image file.

An example of frontispiece encoding:

```
<front>
 <div type="frontispiece">
  <figure>
   <head>Sojourner Truth.</head>
   <figDesc>Woodcut of Sojourner Truth.</figDesc>
   <graphic
```

```
        url="http://docsouth.unc.edu/neh/truth50/frontis.html"
        scale="0.5"/>
  </figure>
 </div>
 <ref target="Etc">...</ref>
</front>
```

Narrative of Sojourner Truth, a Northern Slave, Emancipated from Bodily Servitude by the State of New York, in 1828

4.2.4.6.3. Tables of Contents

- Chapter 4.5, Front Matter

You may wish not to include front matter content such as table of contents or lists of illustrations, especially if you plan to automatically generate the contents or lists of illustrations. If you do, however, plan to manually encode the table of contents (or lists of illustrations and similar content), use a <div> (or <div1>) element with an appropriate **@type** attribute (e.g., div type="contents"). Within this division, use the <list> element to mark up the table of contents, list of illustrations, etc. Each list item should have a <ptr> or <ref> element with a **@target** attribute referencing an **@xml:id** attribute on the <pb> or on the <div> (or <div1>) of the referenced page or section. Use <ref> if you wish to transcribe page numbers in the table of contents; use <ptr> if you do not.

4.2.4.6.4. Notes

Use the <note> element to encode the text of a margin note, footnote, endnote, or other note found in the source document. This element may be used for encoding notes "inline" at the point of reference (such as where a superscript number appears), as in the Alger Hiss example below. In the case of conversion from OCR and from some born-digital source documents, this will require manual intervention to move the text of the note to the place of reference.

Alternatively, the <note> element may encode the text of the note at the point it occurs on the page or at another point convenient when converting from a born-digital source document, such as at the end of the containing <div> (or <div1>) or in a special <div> (or <div1>) element within <back>. The point of reference should be encoded using a <ref> or <ptr> element, as in 3.6 Simple Links and Cross-References. According to this model, the first footnote reference in the Alger Hiss example would be encoded as:

```
<ref target="#n68">68</ref>
```

and the note itself as:

```
<note
  place="bottom"
  anchored="true"
  xml:id="n68"
  n="68">Also before 1947, [...]</note>
```

Marginal notes without reference from the base text should occur at the beginning of the paragraph to which they refer, with `place="margin"`.

Optionally combine notes that extend beyond one page into one <note>.

4.2.4.7. Level 3 Examples

Note that for technical reasons the namespace is not shown in these examples, but it should always be supplied on the root <TEI> element, e.g.: TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="MBFG0236".

4.2.4.7.1. Level 3 Basic Structure: Prose

```
<TEI xml:id="MBFG0236">
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
 <text xml:lang="en">
  <front>
   <div type="frontispiece">[figure]</div>
   <titlePage>[text]</titlePage>
   <div type="dedication">[text]</div>
   <div type="contents">[text]</div>
  </front>
```

```
  <body>
   <div type="book">
    <head>[book title]</head>
    <div type="chapter">[text]</div>
    <div type="chapter">[text]</div>
    <div type="chapter">[text]</div>
    <div type="chapter">[text]</div>
    <div type="chapter">[text]</div>
   </div>
  </body>
  <back>
   <div type="appendix">[text]</div>
   <div type="index">[text]</div>
  </back>
 </text>
</TEI>
```

4.2.4.7.2. Level 3 Basic Structure: Verse

```
<TEI xml:id="VAA2383">
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
 <text xml:lang="en">
  <front>
   <titlePage>[text]</titlePage>
   <div type="dedication">[text]</div>
   <div type="contents">[text]</div>
  </front>
  <body>
   <div type="book">
    <head>[book title]</head>
    <div type="part">
     <head>[section title]</head>
     <div type="poem">
      <head>THE DAYS GONE BY.</head>
      <lg>
       <l>O the days gone by! O the days gone by!</l>
       <l>The apples in the orchard, and the pathway through the rye;</l>
       <l>The chirrup of the robin, and the whistle of the quail</l>
       <l>As he piped across the meadows sweet as any nightingale;</l>
       <l>When the bloom was on the clover, and the blue was in the sky,</l>
       <l>And my happy heart brimmed overin the happy days gone by.</l>
      </lg>
      <lg>[lines of poetry]</lg>
      <lg>[lines of poetry]</lg>
      <lg>[lines of poetry]</lg>
     </div>
    </div>
   </div>
  </body>
 </text>
</TEI>
```

4.2.4.7.3. Level 3 Table of Contents

```
<!--target attribute references page break identifier-->
<div type="contents">
 <head>CONTENTS</head>
 <list type="simple">
  <item>I. A Boy and His Dog <ref target="#VAA2383_011" rend="text-align: right">3</ref>
  </item>
  <item>II. Romance <ref target="#VAA2383_020" rend="text-align: right">12</ref>
  </item>
  <item>III. The Costume <ref target="#VAA2383_029" rend="text-align: right">21</ref>
  </item>
  <item>IV. Desperation <ref target="#VAA2383_038" rend="text-align: right">30</ref>
  </item>
  <item>V. The Pageant of the Table Round <ref target="#VAA2383_046" rend="text-align: right">38</ref>
  </item>
 </list>
</div>
```

4.2.4.7.4. Level 3 Chapter with Letter

```
<div type="chapter">
 <pb xml:id="VAA2383_126" n="118"/>
 <head type="main">CHAPTER XIV</head>
 <head type="subtitle">MAURICE LEVY'S CONSTITUTION</head>
 <p>
  <hi rend="font-weight: bold">L</hi>O, SAM!" said Maurice cautiously. "What you
   doin'?"</p>
 <p>Penrod at that instant had a singular experiencean intellectual shock like a flash
   of fire in the brain. Sitting in darkness, a great light flooded him with wild
   brilliance. He gasped!</p>
<!--Text removed from example-->
 <p>"What you doin'?" asked Maurice for the third time, Sam Williams not having decided
   upon a reply.</p>
 <pb xml:id="VAA2383_127" n="119"/>
 <p>It was Penrod who answered.</p>
 <p>"Drinkin' lickrish water," he said simply, and wiped his mouth with such delicious
   enjoyment that Sam's jaded thirst was instantly stimulated. He took the bottle
   eagerly from Penrod.</p>
 <p>"A-a-h!" exclaimed Penrod, smacking his lips. "That was a good un!"</p>
<!--Text removed from example-->
 <p>Penrod uttered some muffled words and then waved both armseither in response or as
   an expression of his condition of mind; it may have been a gesture of despair. How
   much intention there was in this actobviously so rash, considering the position he
   occupiedit is impossible to say. Undeniably there must remain a suspicion of
   deliberate purpose.</p>
<!--Text removed from example-->
<pb xml:id="VAA2383_138" n="130"/>
 <p>The damsel curtsied again and handed him the following communication, addressed to
   herself: </p>
 <floatingText>
  <body>
   <div type="letter">
    <p>"Dear madam Please excuse me from dancing the cotilo with you this afternoon
        as I have fell off the barn</p>
    <p>"Sincerly yours<lb/> "<hi rend="font-variant: small-caps">Penrod
        Schofield</hi>." </p>
   </div>
  </body>
 </floatingText>
</div>
```

4.2.4.7.5. Level 3 Alger Hiss document

```
<TEI xml:id="someid">
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
 <text xml:lang="en">
  <body>
   <div1>
   <pb n="113" facs="00000001.tif"/>
   <head>POINT VIII. BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S CONVICTION
        SHOULD BE VACATED; ALTERNATIVELY, DISCOVERY AND A HEARING SHOULD BE
        ORDERED.</head>
    <p>The nature and extent of surveillance of Hiss, his family and associates was
        not known at the time of trial by the defense. Even now, with the release of
        some of the govern ment documents concerning FBI investigative techniques
        regarding Hiss, the full extent of surveillance -- wiretapping, mail open
        ings, mail covers, physical surveillance, and other intrusive techniques -- is
        still not 'clear. Nevertheless, it is apparent that information gathered
        through the exploitation of unlawful wiretaps and other illegal surveillance
        was used at trial and consequently the conviction must be reversed.
        Alternatively, further discovery and a hearing is essential to a fair deter
        mination regarding these issues.</p>
    <p>FBI surveillance of Hiss began in earnest in 1941 with the institution of a
        mail cover on his incoming correspondence at his home in connection with an
        FBI investigation of possible Hatch Act violations. CN Ex. 98A. Another mail
        cover was placed <pb n="114" facs="00000002.tif"/> on the Hiss mail in 1945,
        and at the same time the FBI obtained toll call records from the Hiss
        residence Telephone for the years 1943 and 1944 as well. CN Ex. 99. In
        September, 1945, the FBI intercepted telegrams to Hiss as well. CN Ex.
        100.</p>
    <p>In late November, 1945, FBI surveillance of the Hiss residence in Washington,
        D.C., escalated. For the third time, a mail cover was instituted beginning on
```

```
          November 28, 1945, which was continued at least until 1946. CN Ex. 101 at p.
          70; CN Ex. 102. Continuous physical surveillance of Hiss was begun as well. CN
          Ex. 101 at p. 72. Although this twenty-four-hour surveillance was discontinued
          on December 14, 1945, physical surveillance was conducted frequently at
          various times until September, 1947. <note place="bottom" anchored="true" n="68">Also before 1947, a letter
from Priscilla Hiss addressed to her son,
          Timothy Hobson, was intercepted and its contents read. CN Ex. 100A at p.
          167. In approximately March, 1947, a letter from a Michael Greenberg
          addressed to petitioner re garding an application for employment with the
          United Nations was also intercepted, in a manner not revealed by the docu
          ments. CN Ex. 100B</note> CN Ex. 102; CN Ex. 103.</p>
      <p>The most intrusive invasion of petitioner's rights <pb n="115" facs="00000003.tif"/> occurred from December 13,
1945 until the Hisses moved
          from Washington, D.C. to New York City on September 13, 1947. A "technical
          surveillance," -- a wiretap -- was placed on the Hiss telephone at their
          residence on P Street-in Washington, D.C. The logs of this surveillance
          constitute twenty-nine volumes of FBI serials and are roughly 2,500 pages in
          length, in which an enormous amount of information concerning the Hisses' per
          sonal lives, relationships with friends and associates, and habits is
          recorded.</p>
      <p>The wiretap was installed following FBI Director Hoover's application to the
          Attorney General for authorization, <note place="bottom" anchored="true" n="69">Hoover's initial request was
answered by a note reques ting
             information on Hiss. CN Ex. 104. Additional information was furnished by
             letter dated November 30, 1945. CN Ex. 105.</note> although no written
          authorization appears in the documents released to Hiss. The purpose of the
          application was to gather information regarding Hiss' alleged contacts with
          Soviet espionage agents and communists in government service, general
          allegations which had been made by Elizabeth Bentley and Chambers.</p>
      <p>As one would expect, the interception of every telephone</p>
     </div1>
   </body>
  </text>
 </TEI>
```

4.2.4.8. Specification

contains a single TEI-in-Libraries level 3 document, comprising a TEI header and a text, the latter represented as either a transcription (in <text>) or a transcription and page images (in <facsimile>), either in isolation or as part of a <teiCorpus> element. The @rend, @rendition, and @xml:space attributes are not permitted on the root TEI element or within the teiHeader element

Note that for technical reasons the namespace is not shown in this example, but it should always be supplied on the root <TEI> element, e.g.: TEI xmlns="http://www.tei-c.org/ns/1.0".

```
<TEI>
 <teiHeader xml:lang="en">
  <fileDesc>
   <titleStmt>
    <title>A Short Level 3 Document</title>
   </titleStmt>
   <publicationStmt>
    <p>Only published as an example.</p>
   </publicationStmt>
   <sourceDesc>
    <biblStruct>
     <monogr>
      <title>The Princess Bride</title>
      <title type="sub">S. Morgenstern's Classic Tale of True Love and High Adventure</title>
      <imprint>
       <publisher>Harcourt Brace Jovanovich</publisher>
       <date when="1973"/>
      </imprint>
     </monogr>
     <idno type="isbn-10">0-345-41826-3</idno>
    </biblStruct>
   </sourceDesc>
  </fileDesc>
 </teiHeader>
 <text xml:lang="en">
  <body>
   <div type="chapter" n="1">
    <head>The Bride</head>
    <p>The year that Buttercup was born, the most beautiful woman <lb/>in the
        world …</p>
      <!-- rest of chapter 1 -->
   </div>
   <div type="chapter" n="2">
    <head>The Groom</head>
    <note>
```

```
     <p>This is my first major excision. Chapter One, The Bride, is almost
     <lb/>in its entirety about the bride. …</p>
   </note>
      <!-- rest of chapter 2 -->
  </div>
      <!-- rest of the book -->
  </body>
 </text>
</TEI>
```

This element is required. The TEI namespace should be specified on this element, e.g. TEI xmlns="http://www.tei-c.org/ns/1.0".

in a bibliographic reference, contains the name (typically encoded as <persName> or <orgName>) of the author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.

```
<author>
 <orgName>British Broadcasting Corporation</orgName>
</author>
```

```
<author>
 <persName ref="persons.xml#mdalmau.cny">Michelle Dalmau</persName>
</author>
```

```
<author>
 <name>Gibson, Matthew</name>
</author>
```

```
<author>anonymous</author>
```

```
<author>unknown</author>
```

Particularly where cataloguing is likely to be based on the content of the header, it is advisable to use a generally recognized name authority file to supply the content for this element. The attributes **@key** or **@ref** may also be used to reference canonical information about the author(s) intended from any appropriate authority, such as a library catalogue or online resource.

In the case of a broadcast, use this element for the name of the company or network responsible for making the broadcast.

Where an author is unknown or unspecified, this element may contain text such as *Unknown* or *Anonymous*.

```
<back>
 <div1 type="appendix">
  <head>The Golden Dream or, the Ingenuous Confession</head>
  <ab>To shew the Depravity of human Nature </ab>
 </div1>
 <div1 type="epistle">
  <head>A letter from the Printer, which he desires may be inserted</head>
  <ab>Sir. I have done with your Copy, so you may return it to the Vatican, if you
     please</ab>
 </div1>
 <div1 type="advert">
  <head>The Books usually read by the Scholars of Mrs Two-Shoes are these and are
     sold at Mr Newbery's at the Bible and Sun in St Paul's Church-yard.</head>
  <ab> The Christmas Box, Price 1d. The History of Giles Gingerbread, 1d. A
     Curious Collection of Travels, selected from the Writers of all Nations, 10
     Vol, Pr. bound 1l. </ab>
 </div1>
 <div1 type="advert">
  <head>By the KING's Royal Patent, Are sold by J. NEWBERY, at the Bible and Sun
     in St. Paul's Church-Yard.</head>
  <ab> Dr. James's Powders for Fevers, the Small-Pox, Measles, Colds, &amp;c. 2s.
     6d Dr. Hooper's Female Pills, 1s. </ab>
 </div1>
</back>
```

```
<body>
 <head>LA FOREST NVPTIALE,</head>
 <p>Où eſt repreſentee vne varieté bigarree, non mois eſmerveillable que plaiſan-
    te, de diuers mariages, ſelon qu'ils ſont obſerueZ &amp; pratiqueZ par pluſieurs
    peuples &amp; nations eſtranges. Auec la maniere de policier, regir, gouuer- ner
```

```
      &amp; adminiſtrer leur famille.</p>
   <div type="chapter">
    <head>Les Romains. CHAPITRE I.</head>
    <p>Encores que ie ne veuille me formali- ſer contre le droict Romain, neātmoins
       puis que le forma- litez, qui eſtoient an- ciennement gardees au nopces
       Romaines, ſont maintenant miſes hors d'vſages &amp; pratique, ie ne ſe- A
    <pb/> ray point de difficulté d'emprunter des anciens autheurs ce qui appar-
       <!-- ... -->
    </p>
   </div>
  </body>
```

At level 3, the content of <body> may contain only the <head>, <l>, <lg>, <p>, <figure>, <floatingText>, <pb>, <lb>, <note>, and either <div> or <div1> elements.

Use of the **@type** attribute of <div> is recommended

At level 3, <div> may only contain the <head>, <l>, <lg>, <p>, <figure>, <floatingText>, <pb>, <lb>, <note>, and <div> elements.

section section section section section section section contains the name (typically encoded as <persName> or <orgName>) of an individual, institution, or organization acting as editor.

```
  <editor>
   <persName ref="names.xml#khawkins.tvt">Kevin Hawkins</persName>
  </editor>
```

```
  <front>
   <div type="dedication">
    <p>To our three selves</p>
   </div>
   <div type="preface">
    <head>Author's Note</head>
    <p>All the characters in this book are purely imaginary, and if the author has
       used names that may suggest a reference to living persons she has done so
       inadvertently. ...</p>
   </div>
  </front>
```

contains the heading of a division (for example the title of a section), figure, list, line group, or table.

4.2.5. Level 4: Basic Content Analysis

4.2.5.1. Reference

- Chapter 3.3 Highlighting and Quotation, P5 Guidelines
- Chapter 3.4 Simple Editorial Changes, P5 Guidelines
- Chapter 3.5 Names, Numbers, Dates, Abbreviations, and Addresses, P5 Guidelines
- Chapter 3.12.1 Core Tags for Verse, P5 Guidelines
- Chapter 3.12.2 Core Tags for Drama, P5 Guidelines
- Chapter 4, Default Text Structure, P5 Guidelines
- Chapter 6, Verse, P5 Guidelines
- Chapter 7.1.4 Performance Texts: castLists and castItems, P5 Guidelines
- Chapter 7.2.1 Performance Texts: Major Structural Divisions, P5 Guidelines
- Chapter 7.2.2 Performance Texts: Speeches and Speakers, P5 Guidelines
- Chapter 7.2.3 Performance Texts: Stage Directions, P5 Guidelines
- Chapter 13, Names, Dates, People, and Places, P5 Guidelines
- Chapter 14, Tables, Formulæ, and Graphics, P5 Guidelines
- Chapter 14.3 Specific Elements for Graphic Images, P5 Guidelines
- Chapter 16, Linking, Segmentation, and Alignment, P5 Guidelines

Note that this is intended to be a 'TEI conformant' customization, per P5 section 23.3.

4.2.5.2. Purpose

To create text that can stand alone as electronic text, identifies hierarchy and typography, specifies function of textual and structural elements, and describes the nature of the content and not merely its appearance. This level is not meant to encode or identify all structural, semantic, or bibliographic features of the text.

4.2.5.3. Rationale

Greater description of function and content allows for:

- flexibility of display and delivery
- sophisticated searching within specified textual and structural elements
- combining the broadest range of uses and audiences

Level 4 texts contain elements and attributes that describe content, not just appearance, of the text. Texts encoded at Level 4 are able to stand alone without page images in order for them to be read by students, scholars, and general readers, and the encoding of content allows these texts to work effectively with screen readers and other applications that rely on the structure of a text, not just its appearance.

Finally, functionally accurate encoding in Level 4 texts allows them to be searched or displayed in sophisticated ways. For example, a searcher could limit his or her search in a dramatic text to stage directions or in a verse text to only first lines. In a political tract published by subscription, a search could be confined to names that appear in lists, thus limiting a search to names of people who subscribed to a particular volume. This ability to limit searches becomes more significant as textbases become larger, and thus is of great importance to the library community as it attempts to build into the initial design and implementation of textbases the features needed to enhance interoperability.

Level 4 is most suitable for projects with the following characteristics

- sophisticated search and retrieval capabilities are desired
- the texts will be used for textual analysis
- extensibility is desired; that is, one desires to keep open the option for a higher level of encoding to be added by the scholarly community at a later date
- the users of the texts may have limited storage or display capabilities

4.2.5.4. Workflow

Text is generated by keyboarding (likely outsourced double keyboarding from page images using TEI Tite) or possibly by correcting OCR text using software that identifies spelling mistakes and consults a log from the OCR software to find regions of uncertainty in the OCR text. If converting from TEI Tite, minimal additional markup should be added, as discussed in Appendix A of TEI Tite.

4.2.5.5. Element Recommendations for Level 4

Use all elements specified in Levels 1, 2, and 3 *except <ab>, plus elements in the following table*. Note that some of these elements are defined in Level 3 as well, but their use in Level 4 is more strict.

<titlePage> and appropriate child elements	Recommended.
<group>	Recommended to encode a collection of independent texts that are regarded as a single group for processing or other purposes.
<div> *or* <div1>, <div2>, <div3>, etc.	Recommended for encoding a hierarchy of textual divisions. Use as many levels of hierarchy as needed to represent the source document.
<head>	Recommended if headings are present. As in TEI, this element must be the first child of a textual division.
<floatingText>	Recommended when a floating text is identified.
<list> and <item>	Recommended to indicate ordered and unordered list structures.
<table>, <row>, and <cell>	Recommended to indicate table structures.
<hi>	Recommended to indicate change in rendition when a more specific element is not being used; **@rend** attribute is optional.
<opener>, <dateline>, <salute> <closer>, <signed>, <postscript>	Recommended to indicate specific parts of letters.
<castList>, <castItem>, <sp>, <speaker>, and <stage>	Recommended to encode different structures in performance texts (i.e. drama).
<sp> and <speaker>	Recommended to encode oral history interviews.
<epigraph>	Recommended for encoding epigraphs found as front matter
quote rend="___"	Recommended for encoding blockquotes that appear outside the flow of a paragraph. In the **@rend** attribute, give a CSS

	declaration-block (such as padding-left: 0.5in;)
<argument>	Recommended to encode a list of topics sometimes found at the start of a chapter or other textual division.
<trailer>	Recommended to encode a closing title or footer at the end of a division.
<quote>, <said>, <mentioned>, or <soCalled>	Optional.
<emph>, <foreign>, <gloss>, or <term>	Optional.
title type="_"	Optional within the <text> (not the <teiHeader>), especially when text is typographically distinct. Optioanlly use the **@type** attribute with a value as given in the full TEI guidelines except for main titles. (The main value should be used, when appropriate, for <title>s within a TEI header, but is not needed for <title>s elsewhere in a document.)
<ptr> and <ref>	In addition to using to point to notes (as in Level 3), optionally use for identifying cross-references within the text.
<sic>, <corr>, or <choice>	Optionally use to encode errors or typos.
<add>, <del>, <gap>, and <unclear>	Optionally use to encode material that is added, marked for deletion, or is illegible, invisible, or inaudible.
<persName>, <placeName>, <geogName>, and <orgName>	Optionally use to encode personal, place, and organizational names used in a text.
<listName>, <listPlace>, and <listOrg>	Optionally use in support of personal, place, and organizational names normalization and to capture additional information about the names. Should be captured in an external TEI file or database for easier maintenance of names.
<listBibl>	Optionally use in support of bibliographies. Should contain a series of <bibl> elements, which may be further encoded using elements such as <author>, <title>, <publisher>, <biblScope>.

4.2.5.6. General Level 4 Recommendations and Examples

There are many optinoal but not recommended elements at Level 4. While content for many of these elements can be identified within running prose based on changes in typography or use of quotation marks in the source document, they are not always so easily idenitified, or they may occur so often that identification of each instance is impractical. Use only those optional elements that are appropriate for your users' needs and your encoding budget.

- The use of <group> is recommended when you need to encode a body of distinct texts that are grouped together and are regarded as a unit. Most typical examples of such composite texts would be anthologies, collected works of an author, etc. Section 4.3.1 Grouped Texts states,

    The presence of common front matter referring to the whole collection, possibly in addition to front matter relating to each individual text, is a good indication that a given text might usefully be encoded in this way.

- Use <argument> to encode a prefatory list or prose description of the topics usually discovered at the beginning of a chapter. The content within the <argument> element can be presented as a list or as a paragraph:

```
<div type="chapter" n="1">
 <pb xml:id="albert14" n="14"/>
 <head>CHAPTER I.</head>
 <head>CHARLOTTE BROOKS.</head>
 <argument>
  <p>Causes of immorality among colored people – Charlotte Brooks – She is sold South –
     Sunday work.</p>
 </argument>
 <p> ... </p>
</div>
```

Octavia V. Rogers Albert *The House of Bondage, or, Charlotte Brooks and Other Slaves, Original and Life Like, As They Appeared in Their Old Plantation and City Slave Life; Together with Pen-Pictures of the Peculiar Institution, with Sights and Insights into Their New Relations as Freedmen, Freemen, and Citizens.* New York: Hunt & Eaton, 1890.

- The <trailer> element is recommended to encode a heading- or title-like content at the end of a textual division:

```
<body>
```

```
  <head>[book title]</head>
  <div type="chapter" n="1">
   <head>[chapter title]</head>
   <p>[text]</p>
   <trailer>Here ends the Chapter 1.</trailer>
  </div>
  <div type="chapter" n="2">
   <head>[chapter title]</head>
   <p>[text]</p>
   <trailer>Here ends the Chapter 2.</trailer>
  </div>
  <trailer>FINIS.</trailer>
 </body>
```

- Typographically distinct text may be encoded using the following elements:
    - to represent speech, thought, quotation, etc.:
        - quote
        - said
        - mentioned
        - soCalled
    - to represent foreign words or phrases, linguistically emphatic or stressed words or phrases, words regarded as technical terms, etc.:
        - emph
        - foreign (e.g. foreign xml:lang="fr")
        - gloss
        - term
        - title
- Any ambiguous typographically distinct text should be encoded as hi (e.g. hi rend="font-weight: bold"). This element may also be used if the more specific elements above are not used.
- Any of the following three methods may be used to encode errors or typos in original texts:
    - the sic element used alone is optional to indicate errors without correcting them
    - the corr element used alone is optional to provide corrections without indicating the initial error
    - the choice element allows both the apparent error and its editorial correction to be recorded, as in the following examples:

```
  <p>He has no Scruple about Fish; but won't touch a bit of Pork, it being
  <choice>
    <sic>expresly</sic>
    <corr>expressly</corr>
   </choice> forbidden by their Law.</p>
```

Thomas Bluett *Some Memoirs of the Life of Job, the Son of Solomon, the High Priest of Boonda in Africa; Who was a Slave About Two Years in Maryland; and Afterwards Being Brought to England, was Set Free, and Sent to His Native Land in the Year 1734.* London: Printed for R. Ford, 1734. or

```
  <p>4. The art of writing she obtained by her own industry and curiosity, and in so
  short a time that in the year 1765, when she was not more than twelve years of
  <choice>
    <sic>age,she</sic>
    <corr>age, she</corr>
   </choice>
  was capable of writing letters to her friends <pb xml:id="p11" n="11"/> on various
  subjects. She also wrote to several persons in high stations.</p>
```

Abigail Mott, 1766-1851 *Biographical Sketches and Interesting Anecdotes of Persons of Colour. To Which is Added, a Selection of Pieces in Poetry.* New-York: M. Day, 1826.

- The elements <add>, <del>, <unclear>, <gap> may be used to indicate instances when a text (i.e. word or part of it, phrase or part of it) has been added, marked for deletion, or to indicate cases where transcription is difficult (<unclear>) or impossible (<gap>) because the material is illegible, invisible, or inaudible (i.e. while transcribing oral history interviews):

```
  <p>But it is well authenticated by the observation of every one, that <del rend="text-decoration: line-
  through" hand="#JHL">their manner</del>
   <add rend="vertical-align: super" hand="#JHL">this way—i.e.
     the above</add> of writing influences the style of compos. of those who practise it
  considerably, when they grow up to years of manhood; for their productions, <del hand="#JHL" rend="text-
  decoration: line-through">instead</del> far from being terse, argumentative,
  convincing, are without head or tail &amp;amp; are generally an incongruous mass mixed up in the
  most disgusting manner, without divisions or heads &amp;amp; in short without a subject (so to
  speak).</p>
```

Class Composition of J. Horace Lacy [January 1851] 1. Lacy, James Horace 1834-1852

```
<p>But I still hope for &amp;amp; trust in God and I believe he will animate our brave
defenders with a superhuman power and we will yet drive from our soil the hated invaders
whose tread <gap reason="ink blot"/> profanation, but this is an hour to try
men's souls—Fort Donelson has been taken by the enemy. Frank was there and covered
himself with honor but his bravery cost him a wound; he was wounded in the leg slightly—a
flesh wound only, you must not be uneasy.</p>
```

Kimberly Family Personal Correspondence, 1862-1864. Transcript of the manuscript, UNC-Chapel Hill, Southern Historical Collection.

4.2.5.6.1. Level 4 Front and Back Matter

Encode each section of front and back matter as their own textual division. Beyond what is described in the P5 Guidelines, note the following:

Titles pages (recto and verso)
>The use of the <titlePage> element with appropriate child elements describing the major features of most title pages is recommended. The child elements are listed in Section 4.6 "Title Pages". <titlePage> must include the verso if present, divided by pb n="verso"/.

Tables of contents, errata, subscription lists, lists of other titles by the same author, and other such lists'
>must use a <list> with <item>s. For an index, use ref target="_____" to mark up page numbers given in the index, with the value of **@target** referring to the **@xml:id** attribute of the <pb> of the referenced page.

4.2.5.6.2. Level 4 Name Tagging

- Chapter 13.1.1, Linking Names and Their Referents

Names should be encoded using <persName>, <placeName>, <geogName>, and <orgName> elements with the **@ref** or **@key** attribute providing a reference to a <person>, <place>, or <org> element in an external file or database for managing name normalization and compilation of additional information such as biographical or geospatial information. See the discussion of **@ref** and **@key** above for how to choose between them.

If using **@key**, provide a unique internal identifier, such as in a local database.

If using **@ref**, an external TEI file may contain an entry for each name, grouped accordingly under <listPerson>, <listPlace>, and <listOrg>, which is uniquely identified with an **@xml:id** attribute. In such a case the value of the **@ref** attribute in the main TEI document (the transcription of the source document) references the value of the **@xml:id** attribute in the external file. (In the examples below, the external file is named **@context.xml** for 'contextual information' and is in the same directory as the source file, but it may be named anything and placed anywhere that can be referenced by a URI.)

When referencing external files or databases, it is strongly recommended to provide an explanation in the <editorialDecl> section of the TEI header. References to controlled vocabularies and national or local authority files may be signified by a prefix in the **@xml:id** attribute (e.g., tgn_0000000 for the *Getty Thesaurus of Geographic Names*). When referencing a controlled vocabulary be sure to specify this information in the <classDecl> section of the TEI header.

- Place-name tagging example in main TEI document (the transcription of the source document):

```
<p>The first Jews arrived in <placeName ref="context.xml#tgn_7012924">Indianapolis</placeName>
in the middle of the 19th century. Primarily immigrants
from <placeName ref="context.xml#tgn_7000084"> Germany</placeName> and other points in central Europe
(though many had lived elsewhere in the <placeName ref="context.xml#tgn_7012149">United States</placeName>
before they arrived in the city), they were drawn from
throughout the Midwest by the growth of commerce and
rail lines in <placeName ref="context.xml#tgn_7012924">Indianapolis</placeName>.
</p>
```

- In the external file context.xml, for maintaining place name normalization and additional information:

```
<listPlace>
 <place xml:id="tgn_7012924">
  <placeName>
   <settlement type="city">Indianapolis</settlement>
   <region type="state">Indiana</region>
  </placeName>
 </place>
 <place xml:id="tgn_7000084">
  <placeName>
   <country xml:lang="de">Deutschland</country>
  </placeName>
 </place>
 <place xml:id="tgn_7012149">
  <placeName>
```

```
    <country>United States</country>
   </placeName>
  </place>
 </listPlace>
```

- Personal and organizational name tagging example in main TEI document (the transcription of the source document):

```
<p>PRIZE LIBRARY GIFT-Indiana University President
<persName ref="context.xml#lcnaf_82134365">Elvis J.
   Stahr</persName> (right), a former law dean and
practicing attorney, reminisces with Professor of Law
<persName ref="context.xml#lcnaf_00113347">W. Howard
   Mann</persName> as the two inspect some of the nearly
3,000 volumes of <orgName ref="context.xml#lcnaf_79006848">U.S. Supreme
   Court</orgName> records recently transferred to I.U.
from the <orgName ref="context.xml#lcnaf_79109178">Indiana Supreme Court
   Library</orgName>. The collection, dating back to 1925,
is one of the oldest and most complete sets in
existence.</p>
```

- In the external file context.xml, for maintaining personal and organization name normalization and additional information:

```
<listPerson>
 <person xml:id="lcnaf_82134365">
  <persName>
   <surname>Stahr</surname>
   <forename type="first">Elvis</forename>
   <forename type="middle">J.</forename>
  </persName>
  <birth when="1916"/>
 </person>
 <person xml:id="lcnaf_00113347">
  <persName>
   <surname>Mann</surname>
   <forename type="first">W.</forename>
   <forename type="middle">Howard</forename>
  </persName>
 </person>
</listPerson>
<listOrg>
 <org xml:id="lcnaf_79006848">
  <orgName>United States. Supreme Court</orgName>
 </org>
 <org xml:id="lcnaf_79109178">
  <orgName>Indiana. Supreme Court</orgName>
 </org>
</listOrg>
```

- Alternatively, instead of using an external file for the authority data, use the **@key** attribute to point to a unique key in a MySQL table that stores information like county name, FIPS county code, and latitude/longitude values:

```
<p>When Harry Byrd "retired" to his orchards and
Rosemont, his new house outside <placeName key="1498453">Berryville</placeName> in 1930, he was
still an energetic young man with a long political
career ahead of him.</p>
```

4.2.5.6.3. Level 4 Embedded Texts

If the embedded text is more than a short quotation, use <floatingText> even if the instance is still only an excerpt of the embedded text.

Personal letters are a common example of an embedded text. While a collection of letters would use a textual division for each letter, if a letter is quoted as part of a larger text, use <floatingText><body>div1 type="letter" (or <floatingText><body>div type="letter" if using unnumbered textual divisions) with <opener>, <dateline>, <salute>, <signed>, <closer>, <postscript> included as appropriate. For example:

```
<p>She opened and read as follows:</p>
<floatingText>
 <body>
  <div1 type="letter">
   <opener>
    <dateline>AUGUSTA, March 4th, 18—</dateline>
```

```
    <salute>
     <hi rend="font-style: italic">Mrs. A. Mitten:</hi>
    </salute>
   </opener>
  <p>"Having recently understood that you have procured a private teacher, we have
      ventured to stop your advertisement, <hi rend="font-style: italic">though ordered to continue it
        until forbid,</hi> under the impression that you have probably forgotten to have it
      stopped. If, however, we have been misinformed, we will promptly resume the
      publication of it. You will find our account below; which as we are much in want of
      funds, you will oblige us by settling as soon as convenient. Hoping your teacher is
      all that you could desire in one,</p>
  <closer>
    <salute>"We remain, your ob't. serv'ts,</salute>
    <signed>"H—&amp;amp; B—&#x201D;</signed>
  </closer>
  </div1>
 </body>
</floatingText>
```

Augustus Baldwin Longstreet, 1790-1870 *Master William Mitten: or, A Youth of Brilliant Talents, Who Was Ruined by Bad Luck.* Macon, Ga.: Burke, Boykin, 1864.

4.2.5.6.4. Level 4 Drama

Within the front matter (<front>) of a performance text, cast lists must be encoded as <castList>s, with each item in that list encoded as a <castItem>. If desired, each <castItem> may be uniquely identified with an **@xml:id** attribute.

For example,

```
 <front>
  <castList>
   <head>Dramatis Personae</head>
   <castItem xml:id="kllear">LEAR king of Britain</castItem>
   <castItem xml:id="klfrance">KING OF FRANCE</castItem>
   <castItem xml:id="klburgundy">DUKE OF BURGUNDY</castItem>
   <castItem xml:id="klcornwall">DUKE OF CORNWALL</castItem>
   <castItem xml:id="klalbany">DUKE OF ALBANY</castItem>
   <castItem xml:id="klkent">EARL OF KENT</castItem>
   <castItem xml:id="klgloucester">EARL OF GLOUCESTER</castItem>
   <castItem xml:id="kledgar">EDGAR son to Gloucester.</castItem>
   <castItem xml:id="kledmund">EDMUND bastard son to Gloucester.</castItem>
    <!-- ... -->
  </castList>
 </front>
```

Shakespeare's *King Lear*

Within the body of performative texts:

- speeches are encoded as <sp> and speakers identified by the <speaker> element, which is a child of <sp>.
- Stage directions are encoded as <stage> and enclose content describing scenery, stage directions, etc.
- When encoding the actual speech content itself, utilize elements and attributes that correspond to the type of dramatic speech presented (e.g. <p> for prose speech with <lb> to designate a new line in a particular edition of the text or <lg> and <l> to describe dramatic verse structures).
- If normalizing the speaker(s) of a speech is desired, the **@who** attribute of <sp> may be used to refer to the <castItem> of the speaker. When **@who** is used, <speaker> is optional.

```
  <div type="act" n="1">
   <head>Act 1</head>
   <div type="scene" n="1">
    <head>Scene 1</head>
    <stage>King Lear's palace.</stage>
    <stage>Enter KENT, GLOUCESTER, and EDMUND</stage>
    <sp n="1" who="#klkent">
     <speaker>KENT</speaker>
     <p>I thought the king had more affected the Duke of<lb/>
        Albany than Cornwall.</p>
    </sp>
    <sp n="2" who="#klgloucester">
     <speaker>GLOUCESTER</speaker>
     <p>It did always seem so to us: but now, in the<lb/>
        division of the kingdom, it appears not which of<lb/>
        the dukes he values most; for equalities are so<lb/>
        weighed, that curiosity in neither can make choice<lb/>
        of either's moiety.</p>
```

```
    </sp>
    <sp n="3" who="#klkent">
     <speaker>KENT</speaker>
     <p>Is not this your son, my lord?</p>
    </sp>
    <!-- ... -->
   </div>
  </div>
```

4.2.5.6.5. Level 4 Oral History

Speakers in oral history interviews, i.e. interviewee(s) and interviewer(s), may be identified in the <teiHeader> as a list of <author> elements (typically each with a single <persName>) within <fileDesc> / <titleStmt>.

In either method, use an **@xml:id** on the <persName> element to uniquely identify the individual participant:

- The list of an interview's participants can be also listed within the body of the interview (see example below).
- Questions and answers from interviewees and interviewers are encoded as <sp>, with each speaker identified either
    - within <speaker> elements, which are the first child of <sp>, or
    - with a **@who** attribute on <sp>, the value of which poiints to the the <item> for the given speaker in the list of interview participants (by its **@xml:id**), or
    - both.

```
<list type="simple">
 <head>Interview Participants</head>
 <item>
  <persName xml:id="spk1" key="wf" type="interviewee">WILLIAM C. FRIDAY</persName>, interviewee
 </item>
 <item>
  <persName xml:id="spk2" key="wl" type="interviewer">WILLIAM LINK</persName>, interviewer
 </item>
</list>
<!-- ... -->

<sp who="#spk2">
 <speaker n="2">WILLIAM LINK:</speaker>
 <p>Last time we were talking about Frank Porter Graham. And I have a couple of questions
    about Graham, and I wonder if you could clear them up for me. You have mentioned that you
    had worked with him as a student at North Carolina State, had you met him before?</p>
</sp>
<sp who="#spk1">
 <speaker n="1">WILLIAM C. FRIDAY:</speaker>
 <p>No. That budget hearing was the first that I knew of him, of course, but the first time
    that I ever encountered him. I was president of class at N.C. State, and that through me into
    this kind of public adventure. And so I went merrily on downtown and sat there in the budget
    hearing, along with the president of the student body, and some others.</p>
</sp>
```

One possible way to synchronize audio and transcript has been introduced in Oral Histories of the American South, using <milestone> with a **@timestamp** attribute:

```
<milestone
  n="7248"
  unit="empty"
  type="stop"
  timestamp="00:08:54"/>
```

4.2.5.6.6. Level 4 Verse

Use <lg> and <l> as in Level 3. In addition, use the **@rend** attribute to indicate lines that are indented.

For example,

```
<div type="fit" n="1">
 <head>Fit the First: THE LANDING</head>
 <lg type="stanza" n="1">
  <l n="1.1">"Just the place for a Snark!" the Bellman cried,</l>
  <l n="1.2" rend="margin-left: 0.5in">As he landed his crew with care;</l>
```

```
  <l n="1.3">Supporting each man on the top of the tide</l>
  <l n="1.4" rend="margin-left: 0.5in">By a finger entwined in his hair.</l>
 </lg>
 <lg type="stanza" n="2">
  <l n="2.1">"Just the place for a Snark! I have said it twice:</l>
  <l n="2.2" rend="margin-left: 0.5in">That alone should encourage the crew.</l>
  <l n="2.3">Just the place for a Snark! I have said it thrice:</l>
  <l n="2.4" rend="margin-left: 0.5in">What I tell you three times is true."</l>
 </lg>
<!-- ... -->

</div>
```

Lewis Carroll's *The Hunting of the Snark*

4.2.5.6.7. Level 4 Milestones

Instead of using the <milestone> element available in TEI, use ab type="typography". The content of this element is the character(s) or device used to mark the milestone in the source document. For example:

```
<ab type="typography">*****</ab>
```

4.2.5.6.8. Level 4 Alger Hiss document

```
<TEI xml:id="project_document_identifier">
 <teiHeader xml:lang="en"> <!-- header goes here -->
 </teiHeader>
 <text xml:lang="en">
  <body>
   <div1>
    <pb n="113" facs="./pageImages/AH4_0113.jpg"/>
    <head>POINT VIII.</head>
    <head>BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
    <lb/>CONVICTION SHOULD BE VACATED; ALTERNATIVELY,
    <lb/>DISCOVERY AND A HEARING SHOULD BE ORDERED.</head>
    <p>The nature and extent of surveillance of Hiss, his
    <lb/>family and associates was not known at the time of trial by
    <lb/>the defense. Even now, with the release of some of the govern-
    <lb break="no"/>ment documents concerning FBI investigative techniques regarding
    <lb/>Hiss, the full extent of surveillance -- wiretapping, mail open-
    <lb break="no"/>ings, mail covers, physical surveillance, and other intrusive
    <lb/>techniques -- is still not 'clear. Nevertheless, it is apparent
    <lb/>that information gathered through the exploitation of unlawful
    <lb/>wiretaps and other illegal surveillance was used at trial and
    <lb/>consequently the conviction must be reversed. Alternatively,
    <lb/>further discovery and a hearing is essential to a fair deter-
    <lb break="no"/>mination regarding these issues.</p>
    <p>FBI surveillance of Hiss began in earnest in 1941 with
    <lb/>the institution of a mail cover on his incoming correspondence
    <lb/>at his home in connection with an FBI investigation of possible
    <lb/>Hatch Act violations. CN Ex. 98A. Another mail cover was placed
    <pb n="114" facs="./pageImages/AH_0114.jpg"/>
        on the Hiss mail in 1945, and at the same time the FBI obtained
    <lb/>toll call records from the Hiss residence Telephone for the
    <lb/>years 1943 and 1944 as well. CN Ex. 99. In September, 1945,
    <lb/>the FBI intercepted telegrams to Hiss as well. CN Ex. 100.</p>
    <p>In late November, 1945, FBI surveillance of the Hiss
    <lb/>residence in Washington, D.C., escalated. For the third time,
    <lb/>a mail cover was instituted beginning on November 28, 1945,
    <lb/>which was continued at least until 1946. CN Ex. 101 at p. 70;
    <lb/>CN Ex. 102. Continuous physical surveillance of Hiss was begun
    <lb/>as well. CN Ex. 101 at p. 72. Although this twenty-four-hour
    <lb/>surveillance was discontinued on December 14, 1945, physical
    <lb/>surveillance was conducted frequently at various times until
    <lb/>September, 1947.<note place="bottom" anchored="true" n="68">Also
        before 1947, a letter from Priscilla Hiss addressed
     <lb/>to her son, Timothy Hobson, was intercepted and its contents
     <lb/>read. CN Ex. 100A at p. 167. In approximately March, 1947,
     <lb/>a letter from a Michael Greenberg addressed to petitioner re-
     <lb break="no"/>garding an application for employment with the United Nations
     <lb/>was also intercepted, in a manner not revealed by the docu-
     <lb break="no"/>ments. CN Ex. 100B</note> CN Ex. 102; CN Ex. 103.</p>
```

```
    <p>The most intrusive invasion of petitioner's rights
    <pb n="115" facs="./pageImages/AH_0115.jpg"/>
     <lb/>occurred from December 13, 1945 until the Hisses moved from
    <lb/>Washington, D.C. to New York City on September 13, 1947. A
    <soCalled>technical surveillance</soCalled>, -- a wiretap -- was placed on the Hiss
    <lb/>telephone at their residence on P Street-in Washington, D.C.
    <lb/>The logs of this surveillance constitute twenty-nine volumes
    <lb/>of FBI serials and are roughly 2,500 pages in length, in which
    <lb/>an enormous amount of information concerning the Hisses' per-
    <lb break="no"/>sonal lives, relationships with friends and associates, and
    <lb/>habits is recorded.</p>
    <p>The wiretap was installed following FBI Director Hoover's
    <lb/>application to the Attorney General for authorization,
    <note place="bottom" anchored="true" n="69">Hoover's
          initial request was answered by a note reques-
     <lb break="no"/>ting information on Hiss. CN Ex. 104<sic/>. Additional information
     <lb/>was furnished by letter dated November 30, 1945. CN Ex. 105<sic/>.</note>
     <lb/>although no written authorization appears in the documents released to
    <lb/>Hiss. The purpose of the application was to gather information
    <lb/>regarding Hiss' alleged contacts with Soviet espionage agents and
    <lb/>communists in government service, general allegations which had
    <lb/>been made by Elizabeth Bentley and Chambers.</p>
    <p>As one would expect, the interception of every telephone</p>
   </div1>
  </body>
 </text>
</TEI>
```

4.2.5.7. Specification

contains a single TEI-in-Libraries level 4 document, comprising a TEI header and a text, the latter represented as either a transcription (in <text>) or a transcription and page images (in <facsimile>), either in isolation or as part of a <teiCorpus> element. The @rend, @rendition, and @xml:space attributes are not permitted on the root TEI element or within the teiHeader element

Note that for technical reasons the namespace is not shown in this example, but it should always be supplied on the root <TEI> element, e.g.: TEI xmlns="http://www.tei-c.org/ns/1.0".

```
<TEI>
 <teiHeader xml:lang="en">
  <fileDesc>
   <titleStmt>
    <title>A Short Level 4 Document</title>
   </titleStmt>
   <publicationStmt>
    <p>Only published as an example.</p>
   </publicationStmt>
   <sourceDesc>
    <biblStruct>
     <monogr>
      <title>The Princess Bride</title>
      <title type="sub">S. Morgenstern's Classic Tale of True Love and High Adventure</title>
      <imprint>
       <publisher>Harcourt Brace Jovanovich</publisher>
       <date when="1973"/>
      </imprint>
     </monogr>
     <idno type="isbn-10">0-345-41826-3</idno>
    </biblStruct>
   </sourceDesc>
  </fileDesc>
  <encodingDesc>
   <editorialDecl n="4"/>
   <tagsDecl>
    <namespace name="http://www.tei-c.org/ns/1.0">
     <tagUsage gi="div">Unnumbered divs used.</tagUsage>
    </namespace>
   </tagsDecl>
  </encodingDesc>
 </teiHeader>
 <text>
  <body>
   <div type="chapter" n="1" xml:id="Ch1">
    <head>The Bride</head>
    <p>The year that <persName>Buttercup</persName> was born, the most beautiful woman
    <lb/>in the world …</p>
<!-- rest of chapter 1 -->
   </div>
   <div type="chapter" n="2" xml:id="Ch2">
    <head>The Groom</head>
```

```
    <note resp="author">
      <p>This is my first major excision. <ref target="#Ch1">Chapter One, The Bride</ref>, is almost
      <lb/>in its entirety about the bride. …</p>
    </note>
<!-- rest of chapter 2 -->
  </div>
<!-- rest of the book -->
  </body>
 </text>
</TEI>
```

This element is required. The TEI namespace should be specified on this element, e.g. TEI xmlns="http://www.tei-c.org/ns/1.0".
in a bibliographic reference, contains the name (typically encoded as <persName> or <orgName>) of the author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.

```
<author>
 <orgName>British Broadcasting Corporation</orgName>
</author>
```

```
<author>
 <persName ref="persons.xml#mdalmau.cny">Michelle Dalmau</persName>
</author>
```

```
<author>anonymous</author>
```

```
<author>unknown</author>
```

Particularly where cataloguing is likely to be based on the content of the header, it is advisable to use a generally recognized name authority file to supply the content for this element. The attributes **@key** or **@ref** may also be used to reference canonical information about the author(s) intended from any appropriate authority, such as a library catalogue or online resource.

In the case of a broadcast, use this element for the name of the company or network responsible for making the broadcast.

Where an author is unknown or unspecified, this element may contain text such as *unknown* or *anonymous*.

```
<back>
 <div1 type="appendix">
  <head>The Golden Dream or, the Ingenuous Confession</head>
  <p>To shew the Depravity of human Nature </p>
 </div1>
 <div1 type="epistle">
  <head>A letter from the Printer, which he desires may be inserted</head>
  <p>Sir.
     I have done with your Copy, so you may return it to the Vatican, if you please</p>
 </div1>
 <div1 type="advert">
  <head>The Books usually read by the Scholars of
     Mrs Two-Shoes are these and are sold at Mr
     Newbery's at the Bible and Sun in St Paul's
     Church-yard.</head>
  <p> The Christmas Box, Price 1d.
     The History of Giles Gingerbread, 1d.
     A Curious Collection of Travels, selected from the Writers of all Nations,
     10 Vol, Pr. bound 1l.
  </p>
 </div1>
 <div1 type="advert">
  <head>By the KING's Royal Patent,
     Are sold by J. NEWBERY, at the
     Bible and Sun in St. Paul's Church-Yard.</head>
  <p> Dr. James's Powders for Fevers, the Small-Pox, Measles, Colds, &amp;c.
     2s. 6d
     Dr. Hooper's Female Pills, 1s.
  </p>
 </div1>
</back>
```

```
<body>
 <head>LA FOREST
    NVPTIALE,</head>
```

```
  <p>Où eſt repreſentee vne varieté bigarree,
    non mois eſmerveillable que plaiſan-
    te, de diuers mariages, ſelon qu'ils ſont
    obſerueZ &amp; pratiqueZ par pluſieurs
    peuples &amp; nations eſtranges. Auec
    la maniere de policier, regir, gouuer-
    ner &amp; adminiſtrer leur famille.</p>
  <div type="chapter">
   <head>Les Romains. CHAPITRE I.</head>
   <p>Encores que ie ne
      veuille me formali-
      ſer contre le droict
      Romain, neātmoins
      puis que le forma-
      litez, qui eſtoient an-
      ciennement gardees au nopces
      Romaines, ſont maintenant miſes
      hors d'vſages &amp; pratique, ie ne ſe-
      A
   <pb/>
      ray point de difficulté d'emprunter
      des anciens autheurs ce qui appar-
      <!-- ... -->
   </p>
  </div>
 </body>
```

At level 4, the content of <body> may contain only <div> or <div1> elements.
Use of the **@type** attribute of <div> is recommended section

At level 4, <div> may only contain <head>, <p>, <pb>, <note>, and more <div> elements.
section subsection contains the name (typically encoded as <persName> or <orgName>) of an individual, institution, or organization acting as editor.

```
<editor>
 <persName ref="names.xml#khawkins.tvt">Kevin Hawkins</persName>
</editor>
```

```
<front>
 <div type="dedication">
  <p>To our three selves</p>
 </div>
 <div type="preface">
  <head>Author's Note</head>
  <p>All the characters in this book are purely
     imaginary, and if the author has used names that
     may suggest a reference to living persons she has
     done so inadvertently. ...</p>
 </div>
</front>
```

heading contains the heading of a division (for example the title of a section), line group, list, figure, table, argument, or group.

The <head> element is used for headings at all levels; software which treats (e.g.) chapter headings, section headings, and list titles differently must determine the proper processing of a <head> element based on its structural position.

A <head> occurring as the first element of a textual division (i.e., <div>, <divN>, or <lg>) is the title of that chapter or section.

4.2.6. LEVEL 5: Scholarly Encoding Projects

Level 5 texts are those that require substantial human intervention by encoders with subject knowledge. These texts might include encodings of semantic, linguistic, prosodic, or other features well beyond the basic structural elements discussed in Levels 1-4 above. They might also include elements for editorial, critical, or analytical additions; manuscript descriptions; translations; or other textual apparatus. It is impossible to make concrete recommendations for encoding at this level since the scholarly analysis required is usually specific to each project; instead, Level 5 offers the full set of P5 elements as needed.

4.2.6.1. Reference

- Complete P5 Guidelines

4.2.6.2. Purpose

To create deeply analytical encoded texts that might be appropriate for specific research purposes, as part of a scholarly publishing project, or for any other encoding practices in library-based text encoding.

4.2.6.3. Rationale

A significant number of library-based projects engage in high-level analytical text encoding as part of their efforts in digitization, scholarly editing, academic support, or other research. Level 5 is intended to represent that work, which can take advantage of the full richness of the complete TEI Guidelines, while still acknowledging the impact of library-specific practices on encoded text that is created under the auspices of a library.

The specific influences of library practice on a Level-5 encoded text are expressed primarily in adherence to the General Recommendations and TEI Header sections above.

4.2.6.4. Element Recommendations and Examples

Because of the vast range of possibilities for Level-5 encoding, these *Best Practices* have chosen to provide neither a list of recommended elements nor any specific examples for this Level.

Please refer to the TEI Header section above for recommendations for the <teiHeader>, and to the General Recommendations section and the Complete TEI P5 Guidelines for element recommendations and usage examples within the <text>.

5. Acknowledgments

This document is the result of a group of individuals with a range of experience with TEI text encoding, which formed together under the TEI Special Interest Group on Libraries and Digital Library Federation umbrellas. We would like to thank and acknowledge all of those who have given their time and expertise to develop these *Best Practices*.

The individuals who served as editors of this document are:

- Kevin Hawkins, University of Michigan
- Michelle Dalmau, Indiana University

The individuals who have contributed to the writing of this document are:

- Syd Bauman, Brown University
- Michelle Dalmau, Indiana University
- Matthew Gibson, University of Virginia
- Kevin Hawkins, University of Michigan
- Lisa McAulay, University of California, Los Angeles
- Renee McBride, University of North Carolina, Chapel Hill
- Melanie Schlosser, Ohio State University
- Natasha Smith, University of North Carolina, Chapel Hill
- Michael Sperberg-McQueen, Black Mesa Technologies LLC
- Vitus Tang, Stanford University
- Richard Wisneski, Case Western University
- Glen Worthey, Stanford University

The individuals who have contributed to the planning of this document are:

- Syd Bauman, Brown University
- Michelle Dalmau, Indiana University
- Matthew Gibson, University of Virginia
- Kevin Hawkins, University of Michigan
- Lisa McAulay, University of California, Los Angeles
- Chris Powell, University of Michigan
- Andrew Rouner, Washington University in St. Louis
- Melanie Schlosser, Ohio State University
- Natasha Smith, University of North Carolina, Chapel Hill
- Perry Willett, California Digital Library
- Richard Wisneski, Case Western University
- Glen Worthey, Stanford University

The individuals who have contributed complementary tools to this document are:

- Syd Bauman, Brown University: schema specifications and translation to a set of TEI ODD customizations
- Michael Sperberg-McQueen, Black Mesa Technologies LLC: Thutmose stylesheets for MARC-to-TEI header mappings

The individuals who have contributed to copyediting of this document are:

- Susan Lorand, University of Michigan
- Becky Welzenbach, University of Michigan

Lastly, we would like to thank the Digital Library Federation (DLF) for sponsoring two in-person meetings as part of the Spring 2008 Forum in Minneapolis, Minnesota, and the Spring 2009 Forum in Raleigh, North Carolina, in support of our revision work. The DLF also provided teleconferencing

support for our regularly scheduled meetings.

6. Appendix A: Recommended Attributes

The following is a list of attributes recommended for use in TEI documents created according to these *Best Practices*. It includes both attributes explicitly mentioned in these *Best Practices* as well as those that the contributors to these *Best Practices* find likely to be used with the elements mentioned.

- **@anchored**
- **@baseForm**
- **@baseTypes**
- **@calendar**
- **@cause**
- **@cert**
- **@class**
- **@cols**
- **@copyOf**
- **@corresp**
- **@degree**
- **@dur**
- **@dur-iso**
- **@ed**
- **@end**
- **@evaluate**
- **@exclude**
- **@facs**
- **@from**
- **@from-iso**
- **@full**
- **@gi**
- **@given**
- **@group**
- **@hand**
- **@height**
- **@ident**
- **@key**
- **@level**
- **@locus**
- **@marks**
- **@met**
- **@min**
- **@n**
- **@name**
- **@next**
- **@notAfter**
- **@notBefore**
- **@org**
- **@part**
- **@period**
- **@place**
- **@precision**
- **@prev**
- **@quantity**
- **@reason**
- **@ref**
- **@rend**
- **@render**
- **@rendition**
- **@resp**
- **@rhyme**
- **@role**
- **@rows**
- **@sex**
- **@scale**
- **@scheme**
- **@scope**
- **@select**
- **@sort**
- **@subtype**
- **@synch**
- **@target**

- **@targetEnd**
- **@targets**
- **@targFunc**
- **@to**
- **@to-iso**
- **@type**
- **@unit**
- **@uri**
- **@url**
- **@usage**
- **@version**
- **@when**
- **@who**
- **@width**
- **@xml:base**
- **@xml:id**
- **@xml:lang**
- **@xml:space**

7. Appendix B: History of This Document

This document was formerly known as *TEI Text Encoding in Libraries Guidelines for Best Encoding Practices*.

The Text Encoding Initiative Guidelines for Electronic Text Encoding and Interchange (referred to as the *TEI Guidelines*) were first published in 1994 and represent a tremendous achievement in electronic text standards by providing a highly sophisticated structure for encoding electronic text. Digital librarians have benefited greatly from the standardization provided by these guidelines, and the potential for interoperability and long-term preservation of digital collections facilitated by their wide adoption.

In 1998, the Digital Library Federation (DLF) sponsored the TEI and XML in Digital Libraries Workshop at the Library of Congress to discuss the use of the *TEI Guidelines* in libraries for electronic text, and to create a set of best practices for librarians implementing them. From this workshop, three working groups were formed, the members of which represented some of the largest and most mature digital library programs in the U.S.

Group 1 was charged to recommend some best practices for TEI header content and to review the relationship between the Text Encoding Initiative header and MARC. To this end, representatives of the University of Virginia Library and the University of Michigan Library gathered in Ann Arbor in early October 1998 to develop a recommended practice guide. This work was assisted by similar efforts that had taken place in the United Kingdom under the auspices of the Oxford Text Archive the previous year. The section on the header is based on a draft of those recommended practices. It was submitted to various constituencies for comment. In 2008 and 2009, it was heavily revised by Melanie Schlosser, Kevin Hawkins, and other members of the TEI SIG on Libraries.

Group 2 was charged with developing a set of recommendations for libraries using the TEI Guidelines in electronic text encoding. This group included the following representatives from six libraries:

- LeeEllen Friedland, Library of Congress
- Nancy Kushigian, University of California, Davis
- Chris Powell, University of Michigan
- David Seaman, University of Virginia
- Natasha Smith, University of North Carolina, Chapel Hill
- Perry Willett, Indiana University (chair)

At the ALA Midwinter Meeting (January 1999), the DLF task force revised a draft set of best practices, called *TEI Text Encoding in Libraries: Guidelines for Best Practices* (often referred to as *TEI in Libraries Guidelines*). The revised recommendations were circulated to the conference working group in May 1999 and presented at the joint annual meeting of the Association of Computers and the Humanities and Association of Literary and Linguistic Computing in June 1999. Version 1.0 was circulated for comments in August 1999. These guidelines were endorsed by the DLF, and have been used by many digital libraries, including those of the task force members, as a model for their own local best practices. Libraries, museums, and end-users have benefitted from a set of best practices for electronic text in a number of ways, including better interoperability between electronic text collections, better documented practices among digital libraries, and a starting point for discussion of best practices with commercial publishers regarding electronic text creation.

Written in 1998, this first iteration of *TEI in Libraries Guidelines* made no mention of XML, XSLT, or any of the other powerful tools that have now become common parlance and practice in creating digital documents and collections. Based on these important changes in markup technology, it came to the attention of the DLF and members of the original Task Force that the *TEI in Libraries Guidelines* required substantial revision. In 2002, the TEI Consortium published a new edition of the complete *TEI Guidelines* that conformed to XML specifications. In order to remain useful, the *TEI in Libraries Guidelines* had to be updated to reflect these developments.

Furthermore, librarians need more guidance than the original *TEI in Libraries Guidelines* provided. There are many library-specific encoding issues which need to be addressed and documented to ensure consistency. The intention of this document is to provide recommended paths of encoding for these issues.

In addition, these library guidelines have the potential to be much more useful if they can serve as a training document from which librarians can learn about text encoding and addressing particular encoding challenges. To fulfill this role, the guidelines require more examples and detailed explanations, giving documentation of the use of TEI in a library context. Librarians also need a set of standards and best practices for vendors and publishers who

create electronic text for digital libraries, so that these collections adhere to the same archival standards as locally-created electronic text collections. With detailed guidelines that could serve as an encoding specification, librarians might encourage vendors to follow the principles in these standards, to facilitate the long-term preservation of commercially published electronic text collections, and more readily allow for cross-collection searching.

In order to facilitate the evolution of this document, another DLF-sponsored Task Force—some of the representatives of which were on the original Task Force—met on October 24-25, 2003 at the Cosmos Club in Washington, D.C.:

- Richard Gartner, Oxford University Library
- Matthew Gibson, University of Virginia Library
- Kirk Hastings, California Digital Library
- Chris Powell, University of Michigan
- Merrilee Proffitt, RLG
- David Seaman, Digital Library Federation
- Natasha Smith, University of North Carolina, Chapel Hill
- Perry Willett, Indiana University (chair)

These representatives met to revise the original *TEI in Libraries Guidelines* in order that they:

- reflect changes occuring within the text encoding world generally and within the TEI community specifically
- further illuminate the different levels of encoding by offering clearer and more robust examples.

After producing Version 2.0 of the Guidelines, this group (with some changes in membership) met again at the Cosmos Club on February 13-14, 2006. Those in attendance were:

- Syd Bauman, The TEI Consortium
- Richard Gartner, Oxford University Library (by phone)
- Matthew Gibson, University of Virginia (chair)
- Chris Powell, University of Michigan
- Merrilee Proffitt, RLG
- David Seaman, Digital Library Federation
- Natasha Smith, University of North Carolina, Chapel Hill
- Perry Willett, University of Michigan

The group then released Version 2.1 in March 2006.

In April 2008, select members from the TEI Consortium Libraries Special Interest Group (SIG) and the DLF-sponsored TEI Task Force partnered to update the *Best Practices*. The revision was prompted by the release of P5, the newest version of the TEI, and the desire to create a true library-centric customization of the TEI. The group convened for a DLF-sponsored meeting at the Spring Forum in Minneapolis, Minnesota to tackle the revision work. Those in attendance were

- Syd Bauman (Brown University)
- Michelle Dalmau (Indiana University)
- Matthew Gibson (University of Virginia)
- Kevin Hawkins (University of Michigan)
- Lisa McAulay (University of California, Los Angeles)
- Chris Powell (University of Michigan)
- Jenn Riley (Indiana University)
- Andrew Rouner (Washington University in St. Louis)
- Melanie Schlosser (Ohio State University)
- Natasha Smith (University of North Carolina, Chapel Hill)
- Perry Trolard (Washington University in St. Louis)
- Perry Willett (then University of Michigan, now California Digital Library)
- Glen Worthey (Stanford University)

Work continued through conference calls, in which Renee McBride (University of North Carolina, Chapel Hill) and Richard Wisneski (Case Western University) also participated, and at a DLF-sponsored meeting that took place as part of the DLF Spring Forum in Raleigh, North Carolina on May 6, 2009.

In April 2009, a year after the revision work began, the significantly revamped *Best Practices* soon to be known as *Best Practices for TEI in Libraries* (version 3) were disseminated for public comment. At DLF that year, a Birds-of-a-Feather session entitled *TEI Text Encoding in Libraries* was held to gather in-person public feedback. Comments received at the in-person meeting, from the TEILIB-L listserv, through a survey, and by direct email were gathered and prioritized at the DLF meeting. Renee McBride (University of North Carolina, Chapel Hill) agreed to map header elements to MARC elements, and Vitus Tang (Stanford University) provided valuable comments. In addition to addressing most of the comments received, it was resolved that Syd Bauman will generate an ODD specification (One Document Does it All; schema, prose documentation, etc.) for levels 1-4, further ensuring interoperability of texts encoded according to these *Best Practices*.

The revised *Best Practices* contain updated versions of the widely adopted encoding 'levels' — from fully automated conversion to content analysis and scholarly encoding. They also contain a substantially revised section on the TEI header, designed to support interoperability between text collections and the use of complementary metadata schemas such as MARC. They also explore the relationship between METS and TEI and the relationship between these *Best Practices* and the new vendor specification, TEI Tite.

The new *Best Practices* also reflect an organizational shift. Originally authored by the DLF-sponsored TEI Task Force, the current revision work is a partnership between members of the Task Force and the TEI SIG on Libraries. As a result of this partnership, responsibility for the *Best Practices* will

migrate to the SIG, allowing closer work with the TEI Consortium as a whole, and a stronger basis for advocating for the needs of libraries in future TEI releases.

TEI SIG on Libraries. Date: 2011-10
This page is copyrighted