

Internal Encoding Document – Official Document Decisions

TEI Guidelines P5

Version 3.6.0. Last updated on 16th July 2019, revision daa3cc0b9

Abbreviations and Expansions	2
Apparent Errors	2
Atypical Spelling	3
Additions, Deletions, Omissions	3
Bibliographic Citations	4
Dates: general	4
Emphatic Words and Phrases	4
Figures	5
Highlighting	5
Lists	6
Milestone Elements	6
Names	6
Notes	7
Numbers & Measures	7
Pagination	8
Punctuation	8
Quotation	9
Simple Links and Cross-References	9
References	10
Regularization	11
Rendition	11

Abbreviations and Expansions

3.5.5 19th century handwriting makes frequent use of two standard methods of abbreviation: contraction, the omission of letters from the middle of a word, and suspension, omitting letters at the end of the word. When it is clear that the author has deliberately omitted one or more letters, we will treat the word as an abbreviation. We will use <choice> and the <abbr> and <expn> tags for an abbreviation of any sort or to provide the expanded version of the abbreviation.

For example:

```
<choice>
  <abbr>A/c</abbr>
  <expn>Account</expn>
</choice>

<choice>
  <abbr>Fcap</abbr>
  <expn>Foolscap</expn>
</choice>
```

See Document 3.5.5 for Abbreviations we expand (as such) and Abbreviations that we do not expand.

Apparent Errors

3.4.1 When the manuscript is manifestly faulty, and can be confidently determined to be in error, we will record both the correction and the original state of the text. A priority is to present a reading transcription. Thus we will indicate that confidently-perceived errors in the original have been corrected.

<sic>: represent text exactly as it is in the manuscript, although it appears incorrect

<corr>: contains correct form of passage that is apparently erroneous in the text

```
<choice>
  <sic>date's</sic>
  <corr> dates</corr>
</choice>
```

Atypical Spelling

3.4.2 The TEI guidelines recognize that the writing of manuscripts by hand will make extensive use of variant forms of archaic or non-standard spellings (e.g. phonetic spellings or

non-standard spellings often for speed). Atypical spelling should not be confused with abbreviation or expansion.

<orig> (original form): original reading, with no regularization or normalization

<reg> (regularization): contains a reading which has been regularized or normalized

For example:

```
<choice>
  <orig>rul</orig>
  <reg>rule</reg>
</choice>
```

See Document 3.4.2 for Original spellings we regularize (as such) and Original spellings we do not regularize.

Additions, Deletions, Omissions

3.4.3

The manuscript is written in 19th century handwriting style and can be difficult to decipher. We will use the elements <gap> and <unclear> to indicate where there is haziness in the manuscript to the point where it cannot be transcribed. Because of the patchwork construction of some of the manuscript's pages, we will prioritize <gap> and <unclear> and will not use the <add> or elements, as it is difficult to tell where additions were made.

When passages in the text are so illegible that nothing can be read, an omission will be marked with <gap>. We will use @reason to clarify the omission. We will use @resp to identify the encoder.

Example:

Text text text <gap reason="illegible" resp=#RGW/>text text text

When words or phrases are so illegible that the text cannot be transcribed with certainty. will use <unclear>. We will use @resp to identify the encoder.

Sa<unclear resp=#RGW>r</unclear>ah

Bibliographic Citations

3.11 The manuscript provides very few bibliographic references and no full citation information. We will use <bibl> to allow for the most flexibility while preserving fidelity to bibliographic references when they do appear. We will not use advanced bibliographic structures at this time.

Dates: general

3.5.4 Dates that occur in the text will be recorded. Dates will be expressed in the @when attribute. The @when attribute will be used to specify a normalized form for the value of the date in a standard form, e.g. yyyy-mm-dd, as in the following example:

<date when="1891-05-02">May 2nd 1891</date>

Partial dates or times (e.g. 1908) can be expressed in the @when attribute by simply omitting a part of the value supplied.

Emphatic Words and Phrases

3.3.2.2 The <emph> element will be used to mark words or phrases which are linguistically emphatic or stressed. In 19th century handwriting, strokes will occur as ascenders or descenders. In some strokes, ascenders or descenders may be exaggerated above or below the writing line, or may run into or connect to other letters. A stroke may also use an extravagant flourish or backward flourish ("curlicue," curl, swirl, squiggle, or loop). When exaggerated strokes with flourish are used in strategic places, they will be encoded as <emph>.

The particular type of rendition associated with the emphasis will be specified using the @style attribute, as in the following examples:

- Underline (and double underline)
- Flourish (more than simple flourish, extravagant embellishment with "curlicue," curl, swirl, squiggle, or loop): <hi rendition="#flourish">T</hi>
- Tall: <hi style="font-size:large">T</hi>
- Thick (irregular thickness of letters, as if bold)

Figures

3.9 A figure is any sort of drawing, illustration, picture, or diagram. We will use a <figure> element to encoded such graphic images in the text. If the image is accompanied by a header or title, the element <head> will be used to transcribe (or supply) the header or title for the graphic if there is one. We will use <figDesc> to give a simple prose description of the graphic. We will not try to recreate the figure with interior text in our encoding, as the facsimile will likely be provided using the <graphic> tag with the associated @url attribute.

Example:

<figure>

<head>Print Shop Floor Layout</head>

<figDesc>The figure shows a simple sketch of the Townsend shop front room with desks, safes, telephones, windows, and a door to the sales shop.</figDesc>

</figure>

Highlighting

3.3.1 We will use the <hi> element in this pass to record only typographic distinctions not already described by more specific rendition attributes for emphatic words and phrases (@style).

Lists

3.7 The manuscript author makes use of lists for various purposes. The <list> element will be used to mark any kind of list: numbered, lettered, bulleted, or unmarked.

Each item in the list will be encoded as <item> which may be accompanied by a <label>. Lists with titles will be given a <head> element.

Complex lists (greater than 3 columns) will be treated in tabular form.

EXAMPLES!

Milestone Elements

3.10.3 We will use <pb> to represent page breaks, but other milestone elements are not necessary for this manuscript. For more information on how we will use <pb>, see 'Pagination'.

Names

3.5 Names will be encoded using three tags: <persName>, <orgName>, and <placeName>.

There are three ways in which a name can be encoded:

1. If the name of a person is given, we encode with <persName>. William Townsend is encoded as <persName>William Townsend</persName>.
2. If the name of an organization is given, we use <orgName>. As in: <orgName>Sheffield Corporation</orgName>.
3. Finally, if a name refers to a person and an organization, we use <persName> within <orgName>. An example would be <orgName><persName>Thomas Firth</persName> & Sons</orgName>

For more on Organization and Date see Module 13 Documentation.

Notes

3.8 A note is any additional comment within the text that does not fit within the main bulk of the text, what TEI describes as “out of the main textual stream.” All notes are marked as `<note>`, regardless of their placement on the page.

Because notes in this manuscript cannot reasonably be attached to a specific point or span within a text, we will not attempt to transcribe notes at the point of their attachment. Notes will be transcribed at the end of an `<ab>` if there is certainty and at the bottom of the page when no anchor can be reasonably determined.

Numbers & Measures

We will use `<num>` as the simplest way to distinguish numbers from the surrounding text. We will use `<num>` when: 1) the number is seemingly random, 2) we have little specific information about how the number is being used, or 3) when we cannot confidently encode with a stronger option. However, when the number can be encoded as a date or year, street, or monetary value, we will use a stronger encoding.

(Beware potential reference to other page numbers! Cripe...ref! And of these Roman numerals)

We will use `<measure>` as a more robust means of encoding, in that it will give numbers some meaning. Measure contains a word or phrase referring to some quantity of an object or commodity, usually comprising a number, a unit, and a commodity name. Because at this time we cannot confidently encode for type, quantity, unit, or commodity, we will use the global attribute `@ana` to interpret meaning. The values here may be used and it may be the case that multiple values are used.

Values for `@ana`: Finance, Commerce or buying/selling, Business as in expense/income, Money, Pound, Shilling, Pence, Bookbinding, Materials, Dimensions, Paper, Boards, Glue, Leather, Ribbon, Tape, Staff, Wage, Hours, Days, Other?

Examples:

```
<measure ana="business" ana="wage">38</measure>
```

In the case of tables filled with numbers, `@ana` values will be assigned to the table as a whole, not values of individual cells.

From 0041:

```
<table rend="boxed" rows="2" cols="3" ana="business" ana="commerce">
  <row role="label">
    <cell role="label">Cash</cell>
    <cell role="label">Abatements</cell>
```

```

        <cell role="label">Not Paid </cell>
    </row>
    <row role="data">
        <cell role="data">Cash</cell>
        <cell role="data">Abatements</cell>
        <cell role="data">Not Paid </cell>
    </row>

```

Pagination

We will use both <pb> (page break) and <fw> (forme work) to indicate pagination in the manuscript. The <pb> element is used to represent all page breaks and will be used as such with @n for the sequential number and @facs to point to the facsimile. The @facs value corresponds to the filename of the reproduced image. **<pb>s will exist within <ab>, as well as <div>s.**

The purpose of the <fw> element is to record page numbers actually present in the document being encoded, not necessarily to provide a complete or accurate pagination of it. Therefore, we will use the <fw> element with @type to record the number assigned by Townsend to the manuscript pages, since they will always differ from the numerical value of <pb>. Some manuscript pages will not have a <fw> value.

```

<pb facs="0058.jpg" n="34"/>
<fw type="TownsendNum">22</fw>

```

ADD: HOW <PB> WORKS WITHIN <AB> AND <DIVS>

Punctuation

3.2 Punctuation is erratic, sometimes even non-existent. We will likely not run into punctuation marks that are not available in Unicode, but if we do, the <g> element is recommended. We will disregard <pc>. We are not encoding for hyphenation—if a word breaks over two lines, transcribe as a single word. For example, on page 0066, the word “advantage” is broken over two lines, written as “advan-” “-tage”. In our encoding, we are not taking notice of line breaks, so simply encode the word as “advantage”.

Make sure you use the correct unicode character for whatever punctuation is present (to ensure normalization). This means:

- . for period
- ? for question mark
- “ for left quotation mark
- ” for right quotation mark
- ! for exclamation point
- – for dash

There are instances where a thought comes to a full stop, but has no period at the end of it. Silently regularize these instances. The actual text looks like this:

Letters – Works every room and individual see after Customers want or Complaint also the same in sales shop I have erred in leaving the Front Room, I should have been at Call more and made more of Customers by always being in and not running off to dinner at any note untill I had had some one to fill my place well.

We can silently regularize it with full-stops, leaving it as something like this:

Letters – works every room and individual. See after customers want or complaint also the same in sales shop. I have erred in leaving the front room, I should have been at call more and made more of customers by always being in and not running off to dinner. At any note, untill I had had some one to fill my place well.

The best way to decide where a full-stop should be included is by reading a passage aloud, and seeing where you naturally stop.

Quotation

3.3.3 The manuscript does not quote directly from outside texts and does not directly attribute thoughts to external sources. We may find instances in which to use <q> for text in quotation marks. It is unlikely that we find text attributed to an outside source and will not use <quote> in the first pass.

Simple Links and Cross-References

3.6 The <ptr> element will not be used to encode cross-references or links. We will instead use <seg> to provide elements which can be given identifiers for use when aligning or linking to parts of a document.

EXAMPLE

References

3.10 References are links to other location(s) in the document, or to an external one. The most common use for them in this manuscript is to link a term used to either an external glossary or an embedded glossary.

Ditto

Ditto marks (") will be transcribed as the word "ditto" for proper syntax. In order to link the dittoed item to the content being referenced, we will use a <seg> tag pointing to an @xml:id, as shown in the example below. The @xml:id MUST be created before the <seg> can target it.

When working with ditto or implicit ditto marks, the xml:id will take the format of #item_#### where the #### represents the three digit page identifier and the *item* is a suitable lowercase identifier. If the identifier is multiword, only the first word is lowercase with sequential words starting with an uppercase with no spacing between them.

From 0137:

What the ditto mark references:

```
<cell xml:id="fcap_0137">Fcap</cell>
```

Ditto mark:

```
<cell><seg sameAs="#fcap_0137">ditto</seg></cell>
```

Multi-word

```
<cell xml:id="laPost_0137">La Post</cell>
```

Terms

3.3.4 A <term> contains a single-word, multi-word, or symbolic designation which is regarded as a technical term. This element is used with words or phrases identified as terminological in nature. See the glossary for a list of terms and @type attributes.

Example:

```
<term type="bookbinding">boards</term>
```

External Glossary/Terminology

<ref> in glossaries;

<term> has @target

Regularization

We will silently regularize (or decline to record precise information about) a number of textual features or aspects of presentation.

- Missing period (end stop) at the space following a complete phrase considered as such that can be restored with some degree of confidence???
- Unusual use of capital letters, w/o clear importance

Rendition

Rendition supplies information about the appearance of one or more elements in the source text. [Discussion re:Townsend Manuscript--special characters]

In general, information that we do capture includes the following:

- columns
- emphasis

Generally, we are interested in the concept of “renditional distinction” or “renditionally distinct” words or phrases. In particular, we want to encode features that indicate authorial emphasis, words being given deliberate rhetorical emphasis of some sort (emph).