

Playing with the Guidelines: Topics in Text Encoding at the WWP

The *TEI Guidelines* are appropriately intended to be very broad, covering the encoding of all but the most unusual documents for a variety of purposes. The TEI wisely provides a mechanism for extending, or altering, the DTDs, thus allowing for unforeseen needs.

These unforeseen needs have already cropped up frequently since the TEI was first published. The Women Writers Project has often found it necessary to make extensive use of the TEI extension mechanism in order to validly encode features we consider worthy of

being encoded, without resorting to blatant tag abuse.

In this poster session we will present five areas of encoding where we have extended the TEI DTDs.

Renditional Shift for Non-Emphatic Rhetorical Distinction

How would you encode the italics in the following passage from Katharine Chidley's *The Justification of the Independent Churches of Christ* (p. 9) with unextended TEI? Notice that the extent of italics is not a complete paragraph, but more text is in italics than what could perhaps be considered a <quote>.

This example presents a particular problem in denoting renditional change: while it would be possible to encode "You say" as a <hi> and what follows as <quote>, this seems awkward at best, breaking up into pieces that which

rhetorically is a single unit.

The TEI provides several

options for encoding renditional shifts: elements that indicate specific reasons for a shift (e.g., <emph>, <mention>, <name>) and <hi> which declares agnosticism with respect to the reason for the shift. However, at the WWP we have decided it would be useful to divide the set of renditional shifts for which no particular element is applicable into two groups: those which seem to occur simply for visual effect (e.g., as is often seen on title pages); and those which seem to have some meaning, yet not one of

the specific meanings available (e.g., in many texts whenever a proper noun is used as an adjective it is in italics — "the *French* army").

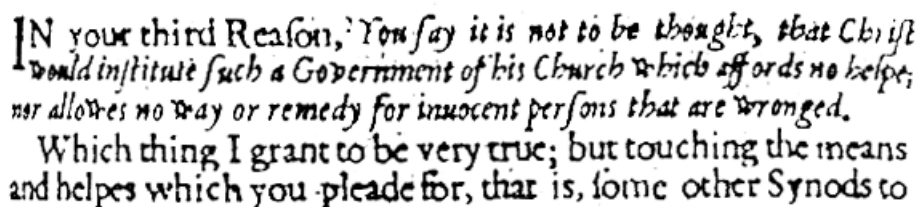
For historical reasons we have chosen to use <hi> to indicate the former group, essentially demoting it from declaring agnosticism to declaring atheism with respect to the rhetorical motivation of the shift. For the latter group we have extended the DTD by adding a new element, <mcr>, which stands for "meaningful change in rendition". Unlike

<hi>, <mcr>

ascribes a rhetorical effect, or meaning, to the renditional

change in the text, without attributing specific meaning, such as the rhetorical weight that <emph> carries, or the change in voice that <quote> indicates, etc.

<mcr> allows us to be precise when marking parts of the text as renditionally distinct. More importantly, <mcr> allows us to describe renditional changes as meaningful, or rhetorically effective, without ascribing value to this meaning or making assumptions about authorial intent which we are not confident to make.



IN your third Reason, You say it is not to be thought, that Christ would institute such a Government of his Church which affords no help, nor allows no way or remedy for innocents that are wronged. Which thing I grant to be very true; but touching the means and helps which you pleade for, that is, some other Synods to

Encoding Figures *sans* Images

How would you encode this figure from *Poetical Essays* by Frances O'Neill? Notice that the caption is in verse.

The unextended TEI content model for `<figure>` is:

```
(head?, p*, figDesc?, text?)
```

where the `<p>` is intended to contain the caption. However, neither `<l>` nor `<lg>` can be a child of `<p>`.

The WWP does not include any reproduction of the image in our encoded file (which would ordinarily go on the entity attribute of the `<figure>` element). Instead we provide a very brief description of the figure within `<figDesc>`. The contents of the `<figDesc>` element will be shown to the user in place of the image.

Our replacement for the content model of `<figure>` is

```
(head?, figDesc, text?, (p | lg)*)
```

Note that the `<figDesc>` is required: this is for local data constraint because we think of the `<figDesc>` as the stand-in for the graphic itself. Thus it also seems reasonable to have the transcription of embedded text and the caption follow the `<figDesc>` rather than straddling it. Our reordering of the elements within `<figure>` places the caption last;



Struggling and hauling both continued long,

we also expanded the content model to allow for verse captions.

We wanted to be able to represent the frequent, simple relationship of caption (or some associated text) to graphic by encoding text-at-the-bottom separately from embedded text, even though we recognize that this is not always an easy line to draw. This is why we preserve the use of `<p>` or `<lg>` rather than having *all* text associated with the graphic encoded within `<text>`.

Multiple Suitors Described Once ... film at 11

In this extract from the castlist of Margaret Cavendish's *A Piece of a Play* two `<castGroup>`s each contain two `<castItem>`s. However, there is only one role description for each `<castGroup>`. The unextended TEI content model for `<castGroup>` is

```
(head?, (castItem | castGroup)+, trailer?)
```

which does not permit a `<roleDesc>` as a child of `<castGroup>`. If the `<roleDesc>` were to be put inside one of the `<castItem>`s it would not appear to apply to the other.

Lord Bear-man,
Sir Puppy Dog-man, } Both Suiters to the Lady Monkey.
Sir Politick Fox,
Monsieur Satyr, } Both Suiters to the Lady Leviret.

Thus we have changed the content model to

```
(  
  head?, roleDesc?,  
  (  
    (label, (castItem | castGroup) )+  
    |  
    (castItem | castGroup)+  
  ),  
  roleDesc?, trailer?  
)
```

Essentially, we allow a single `<roleDesc>` as a child of `<castGroup>` before or after the series of `<castItem>`s. (Note that this content model allows two `<roleDesc>`s; encoding policy restricts us to one.) See the next section for a discussion of the `<label>` added to the content model.

Why is a Castlist Unlike any Other List?

1. **T**he Lord de L'amour.
 2. Sir Thomas Father Love.
 3. Master Comfort, Sir Thomas Father Loves Friend.
 4. Master Charity, the Lord de L'amours Friend.
 5. Adviser, the Lord de L'amours man.
 6. A Justice of Peace.
-
1. The Queen Attention.
 2. The Lady Incontinent, Mistrifs to the Lord de L'amour.
 3. The Lady Mother Love, wife to Sir Thomas Father Love.
 4. The Lady Sanparelle, daughter to Sir Thomas Father, and Lady Mother Love.
 5. The Lady Innocence, the affianced Mistrifs or Wife to the Lord de L'amour.
 6. Passive, the Lady Innocences maid.
 7. Falshood, an informer to maids of the Lady Incontinent.
- Physitians.
 Natural Philosophers, Moral Philosophers, young Students.
 Souldiers, Lovers, Mourners, Virgins, Servants and others.

Because in unextended TEI a <castlist> cannot have <label>s before each <castItem>. However, castlists with labels do exist, as in this one from Margaret Cavendish's *Youths Glory, and Deaths Banquet*. Here is the unextended TEI content model for <castlist>:

```
(
  (%m.divtop;)*,
  (%component;)*,
  (castItem | castGroup)+,
  (%component;)*
)
```

Thus, although a series of <label>s can occur either before or after the main body of <castItem>s, a <label> cannot occur between <castItem>s (or <castGroup>s).

The WWP has changed the DTD to accommodate cases in which all <castItem>s (or <castGroup>s) in a castlist are labelled. Our change parallels the <list> content model in that either all items must have <label>s or none may.

In the following content model, the parameter entity `compsanslabel` is the same as the TEI component, except without <label>.

```
(
  (%m.divtop;)*,
  (%compsanslabel;)*,
  (
    (label, (castItem | castGroup))+
    |
    (castItem | castGroup)+
  ),
  (%compsanslabel;)*
)
```

"Playing with the Guidelines: Topics in Text Encoding at the WWP", presented at the 1999 joint annual conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing, 9-13 June 1999 at the University of Virginia in Charlottesville, Virginia.

By Syd Bauman and Erica Dillon, with assistance from Julie DesJardins and Julia Flanders.

For more information about the Women Writers Project, or to obtain copies of our extensions to the TEI DTD, see <http://www.wwp.brown.edu>.

Updated (minor error corrections, but mostly so I could print it from Classic in OS X) by Syd Bauman in May 2003.

Notes Have Structure, Too!

This endnote from *The Sgelaighe or, a Tale of Old* by Anna Liddiard has internal structure; it is not possible to represent this internal structure using unextended TEI.

The unextended TEI content model for <note> exemplifies the well-known “specialPara” problem. To avoid these problems, and to allow for the encoding of notes that have significant internal structure (e.g., start with a heading and end with a closer or trailer), the WWP completely rewrote the content model for <note>.

In the new model, <note> cannot directly contain PCDATA, so even in notes which consist of a single word or phrase, the content must be

enclosed within a <p> or some other element, as appropriate. This limitation is necessary to eliminate the mixed content problem we had with the TEI content

model. In notes which begin with a member of notetop, the first part of the note (consisting of any notetop elements) should be finished before any element from notecomp

is begun. Thus in a long note beginning with a <ref> and then continuing with some prose explanation, the <p> element should not contain the <ref> element, but rather should be its right sibling. It is conceivable (but we hope unlikely) that a case may occur where the <ref> is syntactically continuous with the following <p> (e.g., “‘Humbler than a duck’ is the author’s first attempt to echo the works of Gaudius”). In such a case, one has to choose between preserving the consistency of the encoding

approach, and preserving the integrity of the paragraph. We continue to discuss this problem.

Note 2.—Page 7.

*“ Still the signal of welcome,
High waved o'er its roof.”*

It was formerly the custom in Ireland for Noblemen and Gentlemen when at home, to erect a *Flag* over their roof, which was universally understood as an invitation to the traveller—who might enter without ceremony, assured of being received with every degree of Hospitality; this laudable custom I am informed is still kept up by a *few—very few respectable* families in the interior, who are as yet *untouched* by the torpedian stroke of *apathetic* Fashion.

The author has from good authority been informed, that the late Dr. Hawkins, Bishop of Raphoe, though he did not exactly keep to the *letter*, adhered to the *spirit* of this custom. There were *certain days, one at least* in each week on which he was “ *at Home*,” not in the meaning of the word as *fashionably* and *foolishly* interpreted, but in the *real* and *sincere* sense which the word originally implied, to all who came within the *focus* of his cheering mansion, and chose to witness and partake of his hospitable board. This custom did not originate ~~from~~ the want of an Inn in the neighbourhood, as there was one very near ~~the~~ Bishop's house. “ *Mine host*,” however, was required on these particular days to mention that the Palace was open to all *Strangers* that came that way.

The Author.

In the following DTD fragment the parameter entity references have been hand resolved. Note that n.byline resolves to “respLine” at the WWP.

```
<!ELEMENT note - - (
  (
    -- notetop -- (ref | head | opener | label)*,
    -- notecomp -- ( bibl | biblFull | biblStruct
      | l | lg | p | sp | cit | q | quote | list | listBibl
      | anchor | seg | note | stage | move | castList)+,
    -- notebot -- (respLine | closer | trailer)*
  )
  |
  ( respLine+ )
)
```

Mixed Content Strikes at specialPara

As briefly mentioned above (in “Notes Have Structure, Too!” on p. 4), the unextended TEI content model for <note> is the specialPara parameter entity reference. With the various base and additional tagsets the WWP uses, this boils down to the following declaration.

```
<!ELEMENT note - O
(
  (
    ( l | lg | p | sp ),
    ( bibl | biblFull | biblStruct | l | lg | p | sp | cit
      | q | quote | label | list | listBibl | note | stage
      | lg1 | camera | caption | move | sound | tech | view
      | castList )*
  )
  |
  ( #PCDATA | abbr | add | addSpan | address | anchor | bibl
    | biblFull | biblStruct | caesura | camera | caption
    | castList | cit | corr | damage | date | dateRange
    | dateStruct | del | delSpan | distinct | emph | expan
    | figure | foreign | formula | fw | gap | gloss | handShift
    | hi | label | link | list | listBibl | measure
    | mentioned | move | name | note | num | orgName | orig
    | ptr | q | quote | ref | reg | restore | rs | seg | sic
    | soCalled | sound | space | stage | supplied | table
    | tech | term | text | time | timeRange | timeStruct
    | title | unclear | view | xptr | xref )*
) >
```

This has the undesirable effect of introducing what is commonly called a “mixed-content” problem, a well known SGML problem which results in ugly use of whitespace in source files.

Thus, noting that the content model for <quote> is exactly the same, we find that

```
<quote>
  <l>Little boxes, on the hillside;</l>
  <l>Little boxes, made of ticky-tacky.</l>
</quote>
```

is not valid SGML, because the whitespace between the start-<quote> tag and the start-<l> tag is interpreted as PCDATA.

However,

```
<quote><l>Little ... hillside;</l><l>Little ... tacky.</l></quote>
```

is valid.

The WWP content model does not suffer this problem. The actual declarations are

```
<!ENTITY % x.notetop '' >
<!ENTITY % m.notetop '%x.notetop; ref | head | opener | label' >
<!ENTITY % x.notecomp '' >
<!ENTITY % m.notecomp '%x.notecomp; bibl | biblFull | biblStruct
  | l | lg | p | sp | cit | q | quote | list | listBibl
  | anchor | seg | note | stage | move | castList' >
<!ENTITY % x.notebot '' >
<!ENTITY % m.notebot '%x.notebot; respLine | closer | trailer' >
<!ELEMENT %n.note; - - (
  ( (%m.notetop;)*,
    (%m.notecomp;)+,
    (%m.notebot;)* )
  |
  ( (%n.byline;)+ )
) >
```

Encodings of Sample Passages

The following are excerpts from valid WWP files. Whitespace has been altered for readability.

Encoding of Sample Passage Demonstrating Use of <mcr>

```
<p rend="first-indent(0)"><hi type="dic">I</hi><hi rend="case(allcaps)">n</hi>
your third Rea&son, <mcr rend="slant(italic)">You &say it is not to be thought,
that <persName>Chri&t</persName>
<lb>would in&stitute &such a Government of his Church which affords no helpe;
<lb>nor allows no way or remedy for innocent per&sons that are wronged.</mcr></p>
<p>Which thing I grant to be very true; but touching the means
<lb>and helps which you pleade for, that is, &some other Synods to <!-- ... --></p>
```

Encoding of Sample Figure

```
<figure>
  <figDesc>An indoor scene of a man and a woman grappling with
    one another; an owl is perched on the woman's shoulder,
    wings spread.</figDesc>
  <lg part="y" type="poem.indeterminate">
    <lg type="indeterminate">
      <quote>
        <l sameAs="101" id="cap01">Struggling and hawling both continued long.</l>
      </quote>
    </lg>
  </lg>
</figure>
<!-- elsewhere: -->
<lg type="poem.indeterminate">
  <lg type="indeterminate">
    <l>She &cream'd, he &wore, and adding blows to
      <lb rend="indent(3)">blows,</l>
    <l>Dragg'd her Scotch arms, &he pull'd his Iri&h
      <lb rend="indent(3)">no&se,</l>
    <l id="101" sameAs="cap01">Struggling and hawling both continued long,</l>
    <l>Both in a rage, both hardy, rough and &trong;</l>
  </lg>
</lg>
```

Encoding of Sample <castGroup>s

```
<castGroup>
  <castItem><role id="RBM">Lord <hi>Bear-man</hi></role>,</castItem>
  <castItem><role id="RDG">Sir <hi>Puppy Dog-man</hi></role>,</castItem>
  <roleDesc>Both Suiters to the <persName>Lady <hi>Monkey</hi></persName>.</roleDesc>
</castGroup>
<castGroup>
  <castItem><role id="RFM">Sir <hi>Politick Fox</hi></role>,</castItem>
  <castItem><role id="RST">Mon&ieur <hi>Satyr</hi></role>,</castItem>
  <roleDesc>Both Suiters to the <persName>Lady
    <hi>Lev<sic corr="eri">ire</sic>t</hi></persName>.</roleDesc>
</castGroup>
```

Encoding of Sample Castlist

```
<castList rend="slant(italic)">
  <castGroup>
    <label>1.</label>
    <castItem>
      <role id="RLDA">
        <hi type="dic">T</hi><hi rend="case(allcaps)">h</hi>e
        Lord de L'amour</role>.</castItem>
    <label>2.</label>
    <castItem>
      <role id="RTFL">Sir Thomas Father Love</role>.</castItem>
    <label>3.</label>
```

```

<castItem>
  <role id="RCOM">Ma&s;iter Comfort</role>,
  <roleDesc>Sir Thomas <lb>Father Loves Friend.</roleDesc></castItem>
<label>4.</label>
<castItem>
  <role id="RCHA">Ma&s;iter Charity</role>,
  <roleDesc>the Lord de <lb>L'amours Friend.</roleDesc></castItem>
<label>5.</label>
<castItem>
  <role id="RADV">Advi&s;ier</role>,
  <roleDesc>the Lord de L'amours man.</roleDesc></castItem>
<label>6.</label>
<castItem>
  <role>A Justice of Peace</role>.</castItem>
</castGroup>
<castGroup>
<label>1.</label>
<castItem>
  <role id="RQOA">The Queen Attention</role>.</castItem>
<label>2.</label>
<castItem>
  <role id="RINC">The Lady Incontinent</role>,
  <roleDesc>Mi&s;tri&s; to <lb>the Lord de L'amour.</roleDesc></castItem>
<label>3.</label>
<castItem>
  <role id="RLML">The Lady Mother Love</role>,
  <roleDesc>wife to <lb>Sir Thomas Father Love.</roleDesc></castItem>
<label>4.</label>
<castItem>
  <role id="RLSP">The Lady Sanparelle</role>,
  <roleDesc>daughter to
    <lb>Sir Thomas Father, and Lady Mo&shy;
    <lb>ther Love.</roleDesc></castItem>
<label>5.</label>
<castItem>
  <role id="RINN">The Lady Innocence</role>,
  <roleDesc>the affianced
    <lb>Mi&s;tri&s; or Wife to the Lord de
    <lb>L'amour.</roleDesc></castItem>
<label>6.</label>
<castItem>
  <role id="RPAS">Pa&s;&s;ive</role>,
  <roleDesc>the Lady Innocences maid.</roleDesc></castItem>
<label>7.</label>
<castItem>
  <role id="RFAL">Fal&s;hood</role>,
  <roleDesc>an informer to maids of the
    <lb>Lady Incontinent.</roleDesc></castItem>
</castGroup>
<castGroup>
<castItem>
  <role>Phy&s;itians</role>.</castItem>
<castItem type="list">
  <role>Natural Philo&s;ophers</role>,
  <role>Moral Phi&shy;<lb>lo&s;ophers</role>,
  <role>young Students</role>.</castItem>
<castItem type="list">
  <role>Souldiers</role>,
  <role>Lovers</role>,
  <role>Mourners</role>,
  <role>Vir&shy;<lb rend="indent(0)">gins</role>,
  <role>Servants</role> and
  <role>others</role>.</castItem>
</castGroup>
</castList>

```

Encoding of Sample Note

```

<note id="note02" type="authorial" target="anch02">
  <label rend="post(&mdash;)">Note 2.</label>
  <ref rend="break(no)">Page 7.
    <quote rend="slant(italic)pre(&ldquo;);post(&rdquo;);break(yes)">Still the
      signal of welcome,
      <lb>High waved o'er its roof.</quote></ref>
  <p rend="first-indent(1)">It was formerly the custom in
    <placeName>Ireland</placeName> for Noblemen and
    <lb>Gentlemen when at home, to erect a <mcr>Flag</mcr> over their roof,
    <!-- ... 5 lines omitted here ... -->
    <lb>by the torpedian stroke of <emph>apathetic</emph> Fashion.</p>
  <p rend="first-indent(1)">The author has from good authority been informed, that
    <lb>the late <persName>Dr. Hawkins, Bishop of Raphoe</persName>, though he did
    <!-- ... 11 lines omitted here ... -->
    <lb><emph>Strangers</emph> that came that way.</p>
  <respLine rend="slant(italic)align(right)">The Author.</respLine></note>

```