

## TEI-XML Audit: Best Practices

Before completing an audit of the TEI-XML files, make sure you have familiarized yourself with the Basics of how TEI-XML works and how oXygen XML Editor works.

The process of auditing double-checks the code for both human error (from the InDesign document) and script errors (from testing). Thus, you need to be generally alert to anything that looks “weird,” be that a mis-applied semantic tag or a clear instance of the script not rendering the correct thing.

---

The best way to get familiar with XML is to simply stare at it in comparison with the corresponding InDesign document. This way, you will learn how the InDesign transformations that we perform for export appear in TEI.

Ex. Compare this view of the opening of a General Introduction with the InDesign document:

```
273 <div>
274 <p style="bodyfirst" n="1">William Caxton's translations of <hi rend="italic">Paris and
275 Vienne</hi> and <hi rend="italic">Blanchardyn and Eglantine</hi> are among the first
276 romances to be printed in England. Published in 1485 and 1489, respectively, they
277 continued to be popular for more than a century and have considerable literary
278 significance. They are the earliest English versions of French narratives already well
279 known in medieval Europe. Circulating in verse and prose versions in numerous languages,
280 romances of the faithful lovers were among the most popular narratives of the later Middle
281 Ages and Renaissance. The stories' sustained and widespread appeal had much to do with
282 their lively, sympathetic characters and to their distinctive treatments of familiar
283 plots. The romances both edify and entertain, exemplifying ideals of chivalry and
284 steadfast devotion while incorporating adventure, suspense, deception, and humor.
285 Audiences would also have appreciated their realistic depictions of tournaments and
286 battles.</p>
287 <p style="body" n="1"><hi rend="italic">Paris and Vienne</hi> and <hi rend="italic">
288 >Blanchardyn and Eglantine</hi> are important witnesses to the development of English
289 prose style and the genre of romance, being in many ways precursors to the novel. They are
290 unique among the romances that Caxton published, for they are the only independent
```

- ❖ The first paragraph is tagged with the Body First paragraph style in InDesign and appears flush with the left margin. In oXygen, the entire paragraph is wrapped with the tag `<p style="bodyfirst" n="1">`
  - It is important to note that, when elements are being drawn from paragraph styles, a new element will open after every end of paragraph mark. This is why there are three separate opening and closing elements for `<p style="body">` when we have three paragraphs in a row all styled with the Body paragraph style.
- ❖ The titles *Paris and Vienne* and *Blanchardyn and Eglantine* are both tagged with the Italic Character Style in InDesign. In oXygen, those phrases are wrapped with the tag `<hi rend="italic">`

- Elements drawn from character styles will only wrap the specific words that have had those particular character styles applied to them.

For a full list of all the TEI elements we use in our XML files, you can refer to the TEI element dictionary.

---

The structure found within <div> containers will change depending on what type of text document you are looking at.

### **Acknowledgments and Dedications**

<p style="bodyfirst"> <p style="body">

These elements will be the only ones wrapping large chunks of text. You may also see other styling-specific elements based on character styles.

### **Abbreviations Lists**

<list>

<item style="abbreviation-listitem">

We want there to only be one opening/closing <list> element in this XML file, with multiple items. You will also see other styling-specific elements based on character styles.

### **Introductions and Headnotes**

The **majority of elements** in Introductions and Headnotes will be some combination of:

<p style="bodyfirst" n="#"> : Body First Paragraph Style

<p style="body" n="#"> : Body Paragraph Style

- ❖ Each of these styles is numbered to create unique ids for internal linking. These numbers will not be consecutive across multiple types of elements, so do expect the numbering to change between bodyfirst, body, and headings.

<p style="bodysubsection"> : Body Subsection Paragraph Style

<p style="bodysubsection2"> : Body Subsection 2 Paragraph Style

<p style="heading"> : Heading Paragraph Style

<p style="heading2"> : Heading 2 Paragraph Style

<p style="heading4"> : Heading 4 Paragraph Style

<div type="footnotes">

<note type="footnote" n="#"> : Footnote Paragraph Style

<note target="#" type="footnotelink">6</note> – these are footnote indicators

**If there are block quotes**, they will be surrounded by one of two elements:

<quote> if the block quote is a prose / larger paragraph

<quote><lg><l> if the block quote is verse.

**If there are tables:**

They will open with the element <table style="basictable">

- ❖ Then the information from within tables will be arranged by row, with individual cells representing each box of information.
- ❖ Any text within these cells will get one of three elements, according to the paragraph style applied to the text:
  - <p style="horizontal-left">
  - <p style="horizontal-center">
  - <p style="horizontal-right">
- ❖ Tables can be tricky. Make sure to consult the InDesign file as you look at the XML.

**If there are lists**, they will look like:

<list rend="bullets">

<item>

<list rend="nobullets">

<item>

<list rend="numbered">

<item style="indent1">

- N.B.: These lists will not have individual numbers for each item in the display text. Instead, the numbers for the list will be added in later since we have defined the list type as “numbered”
- ❖ Lists can also be tricky. Make sure to consult the InDesign file as you look at the XML.

## Verse

<div type="text">

<lg>: Line Group

<l>: lines within stanzas

<milestone n="#" unit="line"/>: Line Numbers

<milestone n="#" unit="stanza"/>: Stanza Number

<locus type="foliation">fol. 23r</locus>: Foliation

<note target="#1" type="footnotelink">1</note>: Footnote Numbers

<note type="enote-indicator" n="#">N</note> Explanatory Note Indicator

<note type="tnote-indicator" n="#">T</note> Textual Note Indicator

<note type="gloss-indicator" n="#">G</note> Gloss Indicator

<p style="heading"> : Heading Paragraph Style

<p style="heading4">: Heading 4 Paragraph Style

<p style="heading5"> : Heading 5 Paragraph Style

<p style="incipit"> : Incipit

- This should be separate from any <lg>s, but should still have <l> nested within them

<trailer> : Explicit

- This should be separate from any <lg>s, but should still have <l> nested within them

<div type="footnotes">

<date>: Date Character Style

<persName> : Person Name Character Style

<placeName>: Place Character Style

## Drama

<div type="text">

<sp>: units of speech. These loosely match up with our use of Line Group Paragraph Style.

However, if multiple speakers swap within a stanza, our XML will prioritize opening a new <sp> to reflect the speaker change.

<speaker>: Speaker

<milestone n="#" unit="line"/>: Line Numbers

<locus type="foliation">fol. 23r</locus>: Foliation

<note target="#1" type="footnotelink">1</note>: Footnote Numbers

<note type="enote-indicator" n="#">N</note> Explanatory Note Indicator

<note type="tnote-indicator" n="#">T</note> Textual Note Indicator

<note type="gloss-indicator" n="#">G</note> Gloss Indicator

<p style="heading"> : Heading Paragraph Style

<p style="heading5"> : Heading 5 Paragraph Style

<p style="incipit"> : Incipit

- This should be separate from any <lg>s, but should still have <l> nested within them

<trailer> : Explicit

- This should be separate from any <lg>s, but should still have <l> nested within them
- <div type="footnotes">
- <date>: Date Character Style
- <persName> : Person Name Character Style
- <placeName>: Place Character Style

## Prose

<div type="chapter" n="#">

<ab>: ab stands for anonymous blocks; we use this in tandem with the Paragraph Style Prose Paragraph

<ab style="incipit">: Prose Incipit

<head type="rubric">: Prose Rubric

<milestone n="#" unit="prose-textblock"/>

<locus type="foliation">fol. 23r</locus>: Foliation

<note target="#1" type="footnotelink">G</note>: Footnote Numbers

- N.B. For Prose we replace the displayed number with a G to indicate that these links contain glosses

<note type="enote-indicator" n="#">N</note> Explanatory Note Indicator

<note type="tnote-indicator" n="#">T</note> Textual Note Indicator

<div type="footnotes">

<postscript><ab></ab></postscript> : Prose Explicit

- N.B. this is different than Verse and Drama for validation reasons.

<date>: Date Character Style

<persName> : Person Name Character Style

<placeName>: Place Character Style

## Facing Page

Instead of the one <div type="text">, there will be two: one for the original language and one for the translation

First: <div type="versotext">

Second: <div type="rectotext">

Third: <div type="footnotes">

Once inside the divs, the code should follow the conventions laid out earlier for [Verse](#), [Drama](#), and [Prose](#), depending on which text type is set in facing page.

- N.B. Facing page will contain no glosses of any kind. You will not see the following:
  - <note type="gloss-indicator" n="#">G</note> Gloss Indicator
  - <note target="#1" type="footnotelink">1</note>: Footnote Numbers

- In prose specifically: <note target="#1" type="footnotelink">G</note>: Footnote Numbers

## Explanatory Notes

<p style = heading>

<p style = heading4>

<note type="enote" n="#">####<space rend="indent1"/><hi rend="italic">

- Keep an eye out to check that the italics aren't wrapping around the <space> element, otherwise the display will look wrong online

<quote>

<lg><l>

<p style="enote-noindent">

## Textual Notes

<p style = heading>

<p style = heading4>

<note type="tnote" n="#">####<space rend="indent1"/><hi rend="italic">

- Keep an eye out to check that the italics aren't wrapping around the <space> element, otherwise the display will look wrong online

<quote>

<lg><l>

<p style="tnote-noindent">

## Bibliography

<listBibl>

<bibl>

## Specific styling elements

These are often drawn from Character Styles.

Usually they are variations on <hi rend="xxx">

- ❖ Italic
- ❖ Bold
- ❖ Bolditalic
- ❖ Smallcaps

<space rend="indent#"/> represents the number of tabs included in a line.

So if there are two tabs in a line, the code will look like <space rend="indent2"/>

## Particular issues to look out for:

Make note of all these issues in a queries document. At this point (intermediate, pre-launch), **we are NOT changing things by hand during this audit**. This is because this audit stage is for us to make sure things are correct in InDesign and/or correct in the transformation script. Once we have that information, we can make changes in the actual XML.

- ❖ Check for split elements
  - When there should only be one element, but for some reason the export has split it into two. Our transformation script should account for most of that, but it still slips through every so often.
  - This can include:
    - Ex.: see `<ref target="insert-internal-link" type="internal">note </ref><ref target="insert-internal-link" type="internal">29</ref>`
    - `<persName>Jesus</persName><persName> Christ</persName>`
    - `<hi rend="italic">Paris an</hi><hi rend="italic">d Vienne</hi>`
    - N.B. Any styling-specific code splits, like the split italics above, are less important to note. Semantic code splits, such as the split `<ref>`s and split `<persName>`s above, are more important because they will interfere with the data and with the functionality of links.
- ❖ Check that any `<ref>` elements have the appropriate type associated with them.
  - `type="internal"`
  - `type="external"`
    - If external, check that the link contained in the target attribute is correct and complete, and it works when you put it in a browser.
- ❖ Check that milestones and foliation are applied correctly to all line numbers, stanza numbers, foliation, etc.
  - This can be easily done by Ctrl F: locus type="foliation" / milestone and clicking Find All; oXygen shows all instances in a menu and you can easily check to make sure that there is continuous numbering.
  - If you click on one of the results, oXygen will take you to the location of the result in the XML document and highlight that specific instance in blue.
- ❖ Check for any extra spaces at the ends of lines. This can be done fairly easily using Ctrl F / Find All in oXygen
  - For Verse, search "`</l>`"
  - For Enotes / Tnotes, search "`</note>`"
  - For Introductions / Headnotes, search "`</p>`"
  - N.B. It is important to include the space before the ending element.
- ❖ Styling elements should only be applied to only the relevant word/phrase
  - Not applied halfway through a word, or include extra spaces before/after the phrase. Keep an eagle eye out for any spacing that looks a little weird!

- ❖ Keep a careful eye on lists. Generally, we want as few opening <list> elements as possible. If there are headings separating parts of a list, this will interrupt the larger <list> element.
- ❖ If there are block quotes, prose or verse, make sure that any citation information is not by itself in a <quote> or <lg> element
  - Citations such as (lines 45-48) should be included in a larger opening element associated with text.
- ❖ Indicator numbering needs to be consistent. Each kind of indicator will be counted on its own system. Make sure that the indicator number in place in the Text matches up with the corresponding catchphrase number in the Notes document.
  - This needs to be checked carefully for glosses.
  - Spotcheck indicators associated with explanatory notes and textual notes.
  - It can help to have the text document and the notes documents (both TEI and InDesign) open at the same time to check whether there are more indicators than notes or vice versa.
- ❖ Make a note of anything else that looks weird.
  - It can be hard to predict how code goes wonky, so if something doesn't look right to you, mark it on your queries sheet. This includes things that look strange in the code and in InDesign as well as things that look strange in the code but fine in InDesign.