Converting Saint Paul: A new TEI P5 edition of *The Conversion of Saint Paul* using stand-off methodology

James Cummings

Oxford University Computing Services, University of Oxford, UK

Abstract

This article reports on the details behind a poster presented a the Text Encoding Initiative (TEI) Members' Meeting at the University of Maryland, College Park, in November 2007. It looks at the creation of of af scholarly electronic edition of a late-medieval play, *The Conversion of Saint Paul* from Bodleian MS Digby 133 using TEI P5 XML. In addition to exploring various new features available in the TEI P5 Guidelines, it also examines the methodology used to create the text, up-scaling from purely presentation markup to descriptive markup, and how this might simplify the creation of such editions. In an attempt to create an interoperable, flexible, and agile edition, it stores anything not directly related to the transcription of the text in separate files in a stand-off manner. In an attempt to experiment with creating a resource which leverages the advantages of networked editions, it documents the attempt to interoperate with the Middle English Dictionary. Although this first appears to be a failure, it highlights some of the inherent problems in attempting to build editions that are dependent on the resources of others. The article concludes with an urge to text encoders to make more of an effort to share examples of, both good and bad, community practice.

Correspondence:
James Cummings, OUCS, University of Oxford, 13 Banbury Road, Oxford, OX2 6NN, UK.
E-mail:
james.cummings@ oucs.ox.ac.uk

1 Introduction

At the Text Encoding Initiative (TEI) Members' Meeting at the University of Maryland, College Park, in November 2007 a poster, upon which this article is based, was presented about the use of the TEI P5 Guidelines (TEI Consortium, 2007) in creating an edition of a late medieval play, with stand-off markup used for interaction with other resources. This was done partly to highlight some of the new features of TEI P5 whose version 1.0 was declared as stable at that meeting; it was also undertaken to experiment with the use of stand-off markup for creating agile interoperable editions. The creation of editions which are intended to be easily interoperable with other editions and reference resources is only going to become more common in the future, and thus testing out ideas and methodologies in this area is inherently beneficial (Robinson, 2003). The text chosen was *The Conversion of Saint Paul*, which is a late-medieval verse play that survives only in a single witness: Bodleian MS Digby 133. The choice of this text is partly because of my academic background in Medieval Drama, and because in the field of Medieval Studies there is a need for freely-available reliable online editions of such plays. This electronic resource is a work in progress; it was initially completed only as a

transcription, but is gradually developing through the addition of secondary material and links to related resources, to become a more complete scholarly edition (Vanhoutte, 2006).

2 New Additions to TEI P5

As an elected member of the TEI Technical Council for the last few years, it has been my privilege to see (and debate) first-hand many new additions to and revisions of the TEI P5 Guidelines. The TEI P5 Guidelines have been followed in production of this edition; specifically it has used a number of features which are new additions to TEI P5. Some of the aspects of the TEI Guidelines used by the edition that are new or were substantially revised for the release of the TEI P5 are discussed below:

(1) A TEI ODD file was created as a way to produce a customized view of the TEI, constrained to only have those elements necessary for this project. One of the benefits of customizing the TEI for each individual project (rather than using example customizations such as TEI Lite) is that then the project has just those, and only those, elements which it needs for the purpose at hand. Moreover, a project can introduce additional constraints such as closed attribute value lists (e.g. one can control just what values are available for the @type attribute on a particular element). These customizations are expressed as a TEI ODD file (where ODD stands for 'One Document Does-it-all') which stores not only the metalanguage necessary for creating a schema (Relax NG, W3C and DTDs are currently supported) with which to validate document instances, but also project-specific reference documentation which can be generated in a number of languages. This can be achieved either by loading an existing TEI ODD into the TEI's ODD processor 'Roma', or using its web front-end to allow easy creation of a customization. In the case of this edition of *The Conversion of Saint Paul*, a TEI ODD was created using Roma which greatly reduced the number of elements available, and generated Relax NG schemas to validate the document (Cummings, 2007).

(2) The edition of *The Conversion of Saint Paul* from Bodleian MS Digby 133 uses the new <msDesc> element to provide relevant metadata in the form of a manuscript description. The TEI P5 Guidelines include an entirely new Manuscript Description module to assist those working with early documents to provide the more detailed catalogue records that such works sometimes need. In this case information about MS Digby 133 as a whole is recorded, including identifying information, history and physical description. Additional information is given about this play text in particular as it appears in the manuscript. The Manuscript Description module has been developed based on many years experience with previous TEI attempts to record this information, most notably the MASTER TEI P4 Extension. The information recorded in the <msDesc> is then used to generate accompanying descriptions of the manuscript as part of the introduction to the edition.

(3) The TEI P5 Guidelines have introduced a new phrase-level element called <choice>, which groups a number of alternate encodings, for example, in order to represent an alternative in textual encoding of a primary source. The edition makes consistent and in-depth use of the new <choice> structure to provide alternative textual information throughout the text. Specifically, in this highly-abbreviated medieval text, this has been used to provide both abbreviations and expansions so that their appearance can then be toggled in the rendered version. The introduction of the <choice> element stemmed partly from the TEI Council's systematic war on free text-bearing attributes. This sustained battle took place not only to facilitate even more robust datatyping of attribute values (e.g. knowing that a value is indeed a date), but also to provide for the occasional need to represent a non-standard character or glyph in what were previously free text attribute values.

The new TEI element for doing so is <g>, but as markup cannot properly exist inside attribute values, most attributes whose values were formally unconstrained text have now been converted to child elements. Thus, with elements for the transcription of primary sources, like <abbr> and <expan> which used to store the alternative encoding as an attribute are now wrapped in a newly-created <choice> structure so both can still be expressed. This leads to create flexibility in the provision of editions based on differing editorial principles, presenting a best text or simultaneously displaying all possible readings (Robinson, 2000).

(4) Scholars encoding highly-abbreviated documents have long asked for a reliable method for indicating which letters have been supplied inside an expanded abbreviation. Using <supplied> is not really appropriate, and nesting <expan> inside <expan> (a previously recommended method) occasionally led to confusion. Hence the TEI P5 Guidelines introduce two new elements: <am> (abbreviation marker) and <ex> (expanded text). Although there has been some debate in the TEI community concerning best practice in a number of complicated edge cases, the introduction of these elements has been widely supported. In the case of this edition of *The Conversion of Saint Paul*, the use of these elements enables the hiding or subsequent display of abbreviation marks in a diplomatic edition view and italicized rendering of the text added to abbreviations in the expanded view.

(5) Another new addition to the TEI P5 Guidelines is the ability to record and link to digital photographic facsimiles of the text being transcribed via the new <facsimile> element. Indeed, it is possible to use this element to represent a written source entirely in the form of a set of images rather than as encoded text. One of the benefits of the new facsimile-related elements is that one is able to indicate surfaces and multiple zones within the facsimile, and image markup tools already exist to help record, annotate, and display these zones. In the case of this edition, information concerning the online digital photographic surrogates provided by the Bodleian was recorded. This is done, as with many of the new pointing attributes in TEI P5, through the use of a URI datatype attributes. This is useful since the images are located on the Bodleian servers and the conditions of use mean that local copies should not be used in the webpages produced. It is hoped that, when time permits, further referencing of various image zones representing interesting textual features will be done at a later date in addition to perfecting a better facing-page text/image display option.

(6) In many instances attribute values which were of type ID and IDREF(S) in the TEI P4 Guidelines (to facilitate referencing from one element to another) have been replaced by URI datatype pointing mechanisms. Although this may seem like a trivial change to some, it is one of the many important changes in TEI P5. Part of the reason for its importance is that it opens up the easy ability to point not only to IDs (now @xml:id attributes) in the current document, but also to any URL on the internet. Thus when providing a @ref attribute to a <persName> element, the attribute value does not have to point to a <person> element (another new TEI P5 addition) in the current document, but it could point to a different file full of people that a project has assembled. Indeed, the @ref attribute could instead point to a standard reference work like the DNB or Wikipedia if desired. Wherever possible, this edition of *The Conversion of Saint Paul* has used the various new URI-based pointing mechanisms in addition to other datatype constraints newly available in TEI P5.

3 Simplified Creation of Agile Editions

In addition to experimenting with new features as they were added to the TEI P5 Guidelines, the

construction of this edition was used to investigate a number of other issues of interest. For example, the initial edition was created quite rapidly, using a number of methods to simplify certain steps in the creation, partly to experiment with such methodologies. During the creation of this edition sufficient structural markup was added (and in some cases generated via XSLT stylesheets) to enable the edition to be re-purposed or further developed during future short bursts of activity when time allowed. The creation of texts with a fairly fine granularity of structural markup in a main file but with additional resources and markup provided in a stand-off manner in other files creates what could be called 'agile editions'. Such editions allow for a greater degree of flexibility in their later uses, for example this edition of The Conversion of Saint Paul, could easily be re-used for a number of research purposes, such as linguistic research or as a performance text, or annotated further by others (Robinson, 2005).

As the edition was being created in spare moments of time in a succession of airport lounges and trains en route to meetings, the methodology used tried to save time and effort wherever possible, but without sacrificing editorial standards (Sperberg-McQueen, 1996). One of the ways this was done was by using images of the manuscript (Bodleian, MS Digby 133) that already exist on the Bodleian's 'Early Manuscripts at Oxford University' website (http://image.ox.ac.uk). This website provides access to over 80 manuscripts in various colleges and libraries associated with the University of Oxford. Some of the images are now well over a decade old and at times this age is evident in the quality of the images. The images are generously pre-licensed for free personal use for teaching and research. However, the Bodleian images are not licensed to use as local copies in your own web pages (you must link to their server rather than redistribute them). While this is completely understandable it has implications for the hosting of a website which is an edition of a play from one of these manuscripts. It is a benefit to research if all projects, where possible, pre-licence their material for easy and free re-use by others.

As the part of the point of this undertaking was to create an edition quickly, it was decided not to transcribe the manuscript from the Bodleian's images. Instead, an out-of-copyright edition of the text was scanned and OCR'd, and this helped to create a rough base text from which to start (Furnivall, 1866). This text was then carefully corrected and proofread letter-by-letter against the Bodleian's online images while comparing to other editions (e.g. Baker, 1982). It might be reasonable to assume that, in creating an XML edition, one would use an XML editor for this initial step, but instead this was done in OpenOffice (a free and open source office application suite). Doing this meant a diplomatic edition of the text could be easily created with presentational formatting used to represent very basic textual phenomena. For example, an italic font was used to represent letters expanded in an abbreviated word, superscript for letters (including abbreviation marks) which were superscript in the manuscript, and to use brackets to enclose any supplied text. This presentational formatting, and a very few others, were sufficient for the purpose. The benefits of using a word processor for this rather than transcribing directly into XML, were the visual ease and postponing of encoding decisions until the entire text had been initially edited. As the manuscript is clearly written, with only a few palaeographic challenges, transcribing (or indeed, correcting OCR) letter-by-letter against images displayed in another window was a fairly straightforward activity. If the process had been more difficult, rather than just editing a few lines at a time in spare moments, it may not have been completed. This can be a very useful methodology for certain types of data-entry projects. However, it tends to work most successfully in non-collaborative projects where version control and the consistency of the application of presentational markup can be more easily controlled by a single individual as any inconsistencies in the presentational markup or character flagging require significant manual correction in the generated XML.

Once the initial text of The Conversion of Saint Paul had been corrected and completed, it was converted to a very basic TEI XML. The underlying format of OpenOffice documents is a compressed form of XML, and thus transforming from these documents to other formats is fairly easy. In the

```
<lb/>
<lb/>
<lb/>P<hi rend="sup">r</hi><ex>i</ex>mus Miles
<lb/>Unto yo<ex>u</ex><hi rend="sup">r</hi> co<ex>m</ex>mau<ex>n</ex>dme<ex>n</ex>t
I do obeysaunce
<lb/>I wyll ^<hi rend="sup">not</hi> gaynsay nor make delac<ex>y</ex>on
```

Fig. 1 Initial XML

case of TEI, there are OpenOffice to TEI filters already created by Sebastian Rahtz, and these worked well enough. The XML markup was then proofread and those presentational elements were searched-and-replaced with more semantically justified ones. In addition, character-based flagging (such as brackets for supplied material) were also replaced with appropriate TEI P5 elements. This form of up-scaling presentational markup to descriptive markup is a common technique to add more intellectual content to a text. For example, while in OpenOffice the expanded portion of abbreviations in italics, and in the conversion to TEI this was transformed to an <emph> element. Although in an HTML rendering of this element, it might still be desirable to display it in italics, it was easily replaced with the (new to TEI P5) <ex> element. This leaves us with XML that looks like Fig. 1.

Although this method produced XML that would be reasonably useful for those simply wishing to display the text, there are many flaws in it for a well-marked up edition. For example, *The Conversion of Saint Paul* is a verse drama, but no speeches are marked here with TEI <sp> elements, nor are the verse lines marked with <l>, their beginnings are simply indicated with the empty element <lb/>.

During the conversion one could have had the lines come out marked, and perhaps even the speeches had the processing had been written from scratch. However, rather than repeat or rewrite the conversion and some of the manual up-scaling that took place afterwards, these were added using XSLT. While one could simply do this inside oXygen (the XML editor) with a series of search-and-replaces using regular expressions, instead a set of small discrete XSLT scripts was chosen. This gave me a chance to experiment more with XSLT2's xsl:for-each-group which is extremely useful in adding markup based on existing markup, text, or spacing.

In the case of *The Conversion of Saint Paul*, a set of three separate tiny scripts were used which copied all the markup in the document while also (1) grouping speeches into into <sp> elements; (2) grouping verse lines into <l> elements; and (3) marking orthographic words in <seg> elements. The first stylesheet takes advantage of two blank carriage returns in between speeches to change these (now two empty <lb/> elements) into a group of <sp> elements (and also assumes that the first line inside the speech was in fact the <speaker>). The second stylesheet groups the individual <lb/> elements inside these speeches into individual line elements enclosing the text. This stylesheet also changed any of the spaces inside these lines, temporarily, to <space/> elements to assist the next stylesheet. The third stylesheet finishes by encoding text inside the lines as <seg> elements indicating orthographic words, giving each of these a unique (and in this case sequential) @xml:id attribute. It did this by grouping together all those things delimited by the temporary <space/> elements added in the previous stylesheet. Adding in the <space/> element for a few seconds between one stylesheet and the other created a simple solution to preserve the arbitrarily nested mark-up, but could have also been done in a variety of other ways. One of the benefits of doing this is that (unlike with search-and-replace in an editor) there is a permanent record in the form of a stylesheet and intermediate files indicating exactly what you have done for future reference. Experience has shown that leaving as many records as possible of the process undertaken in any methodology pays huge dividends. This and the other stylesheets resulted in the following XML (Fig. 2).

This XML is a significant improvement from what existed previously in that it is more descriptively accurate of an interpretation of a dramatic text and there is greater control over what can be

```
<sp>
   <speaker>
      <seg xml:id="w524">P<hi rend="sup">r</hi><ex>i</ex>mus</seg>
      <seg xml:id="w525">Miles</seg>
   </speaker>
   <l>
      <seg xml:id="w526">Unto</seg>
      <seg xml:id="w527">yo<ex>u</ex><hi rend="sup">r</hi></seg>
      <seg xml:id="w528">co<ex>m</ex>mau<ex>n</ex>dme<ex>n</ex>t</seg>
      <seg xml:id="w529">I</seg>
      <seg xml:id="w530">do</seg>
      <seg xml:id="w531">obeysaunce</seg>
   </l>
   <l><seg xml:id="w532">I</seg>
      <seg xml:id="w533">wyll</seg> ^<seg xml:id="w534">
         <hi rend="sup">not</hi>
      </seg>
      <seg xml:id="w535">gaynsay</seg>
      <seg xml:id="w536">nor</seg>
      <seg xml:id="w537">make</seg>
      <seg xml:id="w538">delac<ex>y</ex>on</seg></l>
   <!-- rest of speech removed -->
</sp>
```

Fig. 2 Up-scaled XML

```
<seg xml:id="w527">yo<ex>u</ex><hi rend="sup">r</hi></seg>
```

Fig. 3 An abbreviated word

done with these elements. However, there are still some structural improvements to be made to this markup. While abbreviations have been expanded, there is no easy way of displaying one or the other reading at a touch of a button in a rendered view. In order to do this the <choice> element can be used to provide both an <abbr> and an <expan> for each word with an abbreviation in it. This was achieved with a fairly straightforward XSLT stylesheet. For example, in the figure above there is the word number w527, the word 'your' which is frequently abbreviated as a 'yo' with a superscript 'r' above it. This was in the XML document as (Fig. 3).

That is, a <seg> element indicating the word boundary, an <ex> element indicating the supplied expanded text, and a <hi> element indicating that the 'r' was originally superscript (which should

probably in hindsight have been marked as an <am> element). To modify this a fairly straightforward XSLT transformation, copying all elements except for those of <seg> which have a descendant <ex> element, can be used. To these elements a new choice structure is introduced, while creating <expan> and <abbr> forms. In XSLT this might look something like (Fig. 4).

This changed the <seg> element into one which has a <choice> element embedded in it (Fig. 5).

This gives multiple options for display: abbreviated or expanded and with or without superscript letters. This transformation has up-scaled the markup even more in that the original (already expanded) form has been preserved alongside an abbreviated version which more closely records what was in the manuscript.

```
<!-- Default rule, copy everything -->
<xsl:template match="@*|node()|comment()|processing-instruction()" priority="-1">
  <xsl:copy><xsl:apply-templates select="@*|node()|comment()|processing-instruction()"/></xsl:copy>
</xsl:template>
<!-- Default rule, do the same but in vanish mode -->
<xsl:template match="@*|node()|comment()|processing-instruction()" priority="-1" mode="vanish">
  <xsl:copy><xsl:apply-templates select="@*|node()|comment()|processing-instruction()"/></xsl:copy>
</xsl:template>
<!-- In vanish mode ignore the <ex> element -->
<xsl:template match="tei:ex" mode="vanish"/>
<!-- Create new <choice> structure -->
<xsl:template match="tei:seg[tei:ex]">
  <seg>
    <xsl:apply-templates select="@*"/>
    <choice>
      <abbr>
        <xsl:apply-templates mode="vanish"/>
      </abbr>
      <expan>
        <xsl:apply-templates/>
      </expan>
    </choice>
  </seg>
</xsl:template>
```

Fig. 4 XSLT to create <choice> element

```
<seg xml:id="w527">
  <choice>
    <abbr>yo<hi rend="sup">r</hi></abbr>
    <expan>yo<ex>u</ex><hi rend="sup">r</hi></expan>
  </choice>
</seg>
```

Fig. 5 An abbreviated word with <choice>

4 Generating Stand-off Markup

Another one of the concerns in undertaking the creation of this edition of *The Conversion of Saint Paul* has been an increasing interest in the creation of new resources through stand-off linking between other multiple resources. As mass digitization projects (providing at most barely structurally-encoded text) seem to be becoming more common, it is interesting to explore how publicly-available resources can be combined in order to form something new and useful that is more than the sum of the parts (Eggert, 2005). This is in many ways the same motivation for the so-called Web 2.0 'mash-ups', which combine more than one resource. In order to experiment with this idea all other markup, except for the basic structural and transcriptional markup already discussed, has been provided in a stand-off manner in separate files.

The textual edition of *The Conversion of Saint Paul* forms just one component of the resource as a whole and, as with all good scholarly editions, it is coupled with additional editorial material (Unsworth, 2002). One of the other components is a generated orthographic word-list. This has been generated from the edition itself by using an XSLT stylesheet which lists alphabetically each distinct word in the text. To be specific each word of the transcript is compared using a deep-equals case-insensitive distinct-values comparison and then output with a unique @xml:id attribute. This 'deep-equals' comparison examined not only the spelling of words, but all of their descendant elements, and thus captured differences in

```
<entry xml:id="let27-w11">
  <form>
    <orth type="hw">your</orth>
    <form n="33" type="orthVar">
      <orth xml:id="let27-w11-vA">Yo<ex>u</ex>
        <hi rend="sup">r</hi>
      </orth>
      <orth xml:id="let27-w11-vB">Yo<ex>u</ex>r</orth>
      <orth xml:id="let27-w11-vC">yo<ex>u</ex><hi rend="sup">r</hi></orth>
    </form>
  </form>
</entry>
```

Fig. 6 An <entry> in the word-list

```
<linkGrp xml:id="let27-w11-lg">
  <link targets="wordlist.xml#let27-w11-vA stp.xml#w326 stp.xml#w851"/>
  <link targets="wordlist.xml#let27-w11-vB stp.xml#w883"/>
  <link
    targets="wordlist.xml#let27-w11-vC stp.xml#w70 stp.xml#w323 stp.xml#w363
    stp.xml#w371 stp.xml#w381 stp.xml#w529 stp.xml#w596 stp.xml#w649
    stp.xml#w789 stp.xml#w878 stp.xml#w1136 stp.xml#w1162 stp.xml#w1167
    stp.xml#w1389 stp.xml#w1450 stp.xml#w1836 stp.xml#w2388 stp.xml#w2426
    stp.xml#w2429 stp.xml#w2700 stp.xml#w2782 stp.xml#w2879 stp.xml#w2890
    stp.xml#w2941 stp.xml#w3375 stp.xml#w4903 stp.xml#w4963 stp.xml#w4969
    stp.xml#w5116 stp.xml#w5120" />
</linkGrp>
```

Fig. 7 A link group citing word-list and word instances in the edition

abbreviation/expansion inside individual words. This means an originally abbreviated word and a word originally spelled out in full become part of the same entry, but the same word which is spelled multiple ways, (for example, the quite common occurrence of them having an extra 'e' at the end) end up as different, though adjacent, entries. Each entry contains a regularized headword, the number of instances of this word in the source document, and a list of all the variant orthographic forms found for that spelling. The form for these entries is that of the TEI dictionaries module for a simple dictionary entry (Fig. 6).

What should be emphasized, as a significant limitation, is that this is not a true word index with an entry for each linguistic word, but only a list of distinct orthographic words which does not contain

the <choice> structure used in the edition. Moreover, while it was also entirely possible to record references to the @xml:id values of each textual instance of these words in the word-list entries, this was not done. The intention was to treat these as separate resources and to explore XSLT transformations to deal with linking the two files together. By comparing the edition and the word-list a third file was automatically created containing TEI <linkGrp> elements for each word; these in turn contained <link> elements for each variant orthography. This provided a link between the entry in the word-list to each of the instances in document text through their @xml:id values. For example (Fig. 7).

This linking file was created with XSLT by comparing the source text to the word-list file. This is

inherently the more difficult way to undertake the work since it could have been generated from the source text at the same time, and by the same stylesheet that created the word-list. As part of the point was to practice linking together disparate resources however, it seemed more realistic to treat the initial files as those which could not be changed. This is true of most web resources; we are given access to read them, but if we want to link them together, it must be done in a hands-off manner where we access the original resources on their own terms. The benefits of linking together the individual instances of words in the source text and entries in the word-list are fairly straightforward. In an HTML rendered view of the text each word is a hyperlink and clicking on it takes you to the corresponding entry in a similarly-rendered view of the word-list. While at the entry (courtesy of information imported from the <link> element in the linking file) you can return directly to any other instance of that word in the source text. This could create a very circular method of exploring the text where a user clicks on a word they are interested in, is taken to the appropriate entry in the word-list, and from there see other instances of it in the document, perhaps exploring those. Providing a listing of each occurrence of that word in the text either at the entry in the word-list or indeed as another file is an optional side benefit which can be generated quite simply from the linking file itself.

5 Standing Off from Interoperability

One of the primary interests from a digital humanities perspective in creating this edition has been the use of stand-off markup to provide interoperable linking with other resources. The links discussed in the previous section, however, are in fact with documents generated solely from the source text (a word-list and a linking file); indeed information that could easily have been stored within the original file. It was thus important that this edition should be linked with a useful resource that was freely-available but which could not be changed.

The most appropriate freely-available resource in this case is the Middle English Dictionary (MED), created by the University of Michigan. As the edition was intended to be created as easily as possible, checking each orthographic words in the edition against the MED and linking directly to the real correct senses was deemed too time-consuming. While that would certainly be possible, it would take significantly more time and be prone to error. Instead the desire was to pass the regularized form of a word to the MED headword search engine during the XSLT transformation that created HTML version of the word-list and incorporate the results into the pre-generated display of the entry for that word in the word list. This would have the effect of including the words which the MED felt might be appropriate in the most suitable place in the edition. This proved impossible to do, however, in this manner from the MED website because their output, despite claiming to be XHTML, was not well-formed. The error (at time of writing) in both search result and entry display pages involves the fairly trivial problem of unescaped ampersands as parameters in their hyperlinks, and some other minor problems, which causes any XSLT transformation on the XHTML (requiring well-formed XML) to fail. There are many possible methods to work around this problem, for example one could retrieve the required HTML pages in advance and easily process them into proper XHTML, but this goes against the spirit of exploring interoperability which was the motivation for using another resource as part of the one being created. Indeed, simple links which search the MED on their site could have been created and would allow users to move to the MED at that point. However, as a good member of the digital medievalist community, a bug report was instead submitted (in this case in person at a conference) so that their resource could be made more interoperable. While this hasn't yet happened, it resulted in the offer of some of the underlying MED working files which list the MED ID, headword, and the variant orthographic forms. A quick XSLT transformation changed this to TEI P5 dictionary entries similar to the edition's own word-list (Fig. 8).

```
<entry xml:id="MED54036">
  <form>
    <orth>yǒur</orth>
    <pos>pron.</pos> Also <orth>youre</orth>, <orth>yowir</orth>,
    <orth>yowur</orth>, <orth>yor</orth>, <orth>yure</orth>, <orth>yeur</orth>,
      <orth>yhure</orth>,
    <!-- and many other <orth> elements left out for space in this example  -->
  </form>
</entry>
```

Fig. 8 A converted MED headword entry

Now, when re-generating the the HTML-rendered version of the word-list all the <orth> elements could be searched to find any which matched, and thus grab the @xml:id attributes on any ancestor <entry> elements. These could then be used to create HTML links directly to the MED from the edition's word-list entries. This bypasses the MED search form entirely, though users can of course return to their site, if desired, to perform more complicated searches. While this is a creative and useful solution, it is disappointing because it works counter to the principle of interoperability in that it necessitated duplication of the workings of the MED basic search rather than using it directly. This should be counted as a failure in a test of interoperability, but sometimes it is the failures which become useful catalysts for change – hence the desire to document it here.

6 Conclusions

The investigation and experimentation undertaken in creating this edition highlight one of the many problems with attempting to build a digital resource which relies on combining existing disparate resources; particularly those which are out of one's control leaving one reliant on whatever the resource creators choose to provide. It is partly this that makes the adherence to open standards important as this will help to enable the possibility of interoperability between such resources. In TEI XML based resources in particular, there has long been a publicly declared willingness to share one's underlying data, but in reality only a limited number of these XML files are exposed alongside the rendered HTML versions (my XML will be available from the moment the edition is publicly announced). The excuses for not making the underlying XML available range from licensing concerns to delicate academic egos (for example a reluctance to show XML with so-called 'tag abuse'). However, as a developing set of community-driven principles, the TEI Guidelines are only as good as the documents which follow them. In order to improve the TEI Guidelines one of the things that is needed are examples of community practice, but this means examples of bad practice as well as good. Through the discovery of our joint errors and misunderstandings of the Guidelines we can help to improve them. People are often embarrassed to show the world the underthings of their XML coding, but people should be urged to show all of their dirty laundry! If anything, the desire to create easily transformed, re-purposable agile editions is inspired by a wish to see more of the underlying XML that we tend to hide away.

References

Baker, D. C., Murphy, J. L., and Hall, Jr. L. B. (1982). (eds) *Late Medieval Religious Plays of Bodleian MSS Digby 133 and E Museo* 160, EETS, 283. Oxford: Oxford University Press.

Cummings, J. (2007). The text encoding initiative and the study of literature. In Siemens, R. and Schreibman, S. (eds), *A Companion to Digital Literary Studies*. Oxford: Blackwell Publishing, pp. 451–476.

Eggert, P. (2005). Text-encoding, theories of the text and the 'work-site'. *Literary and Linguistic Computing*, **20**(4): 425–435.

Furnivall, F. J. (ed.) (1896). *The Digby Plays*, (New Shakespeare Society Publications, 1882), Re-issued for EETS Extra Series LXX, Oxford: Oxford University Press.

Robinson, P. (2000). The one and the many text. *Literary and Linguistic Computing*, **15**(1): 5–14.

Robinson, P. (2003). Where we are with electronic scholarly editions, and where we want to be. *Jahrbuch für Computerphilologie*, **5**(2003): 123–143 <http://www.computerphilologie.lmu.de/jg03/robinson.html> (accessed 10 March 2009).

Robinson, P. (2005). Current issues in making digital editions of medieval texts—or, do electronic scholarly editions have a future? *Digital Medievalist*, 1:1, <http://www.digitalmedievalist.org/journal/1.1/robinson/> (accessed 10 March 2009).

Sperberg-McQueen, C. M. (1996). Textual criticism and the text encoding initiative. Annual Convention of the Modern Language Association, 1994. In Reprinted in Finneran, R. J. (ed.), *The Literary Text in the Digital Age*. Ann Arbor: University of Michigan Press, pp. 37–62.

TEI Consortium. (2007). (eds) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (accessed 10 March 2009).

Unsworth, J. (2002). Electronic Textual Editing and the TEI. *Annual Convention of the Modern Language Association, 2002.* <http://www3.isrl.uiuc.edu/~unsworth/mla-cse.2002.html> (accessed 10 March 2009).

Vanhoutte, E. (2006). Prose fiction and modern manuscripts: limitations and possibilities of text-encoding for electronic editions. In Burnard, L., O'Brien O'Keefe, K., and Unsworth, J. (eds), *Electronic Textual Editing*. New York: Modern Language Association of America, pp. 161–180.