Journal of the Text Encoding Initiative

Issue 6 | December 2013

Selected Papers from the 2012 TEI Conference

Stationers' Register Online: A Case Study of a Byte-Reduced TEI Schema for Digitization (`tei_corset`)

James Cummings and Pip Willcox

Electronic version

URL: <http://journals.openedition.org/jtei/926>

DOI: 10.4000/jtei.926

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

James Cummings and Pip Willcox, « Stationers' Register Online: A Case Study of a Byte-Reduced TEI Schema for Digitization (`tei_corset`) », *Journal of the Text Encoding Initiative* [Online], Issue 6 | December 2013, Online since 22 January 2014, connection on 01 May 2019. URL : <http://journals.openedition.org/jtei/926> ; DOI : 10.4000/jtei.926

*Stationers' Register Online: A Case Study of a Byte-Reduced TEI Schema for Digitization (*tei_corset*)*

James Cummings and Pip Willcox

1. Introduction

- 1 This article examines the technical background to the Stationers' Register Online (SRO) project at the University of Oxford.¹ This pilot SRO project received institutional funding from the Lyell Research Fund to transcribe and digitize the first four volumes of the book-entry Register of the Stationers' Company. This is arguably the most important primary source for the study of the history of the book in Britain, apart from the books themselves. The Stationers' Register was "the Company's own primary means of self-regulation ... in order to prevent unfair competition between its members" (Gadd 2004, 42–43), through which ownership of texts was asserted, disputed, regulated, and monitored from 1577 until 1924. The Register survives intact in two series now held in Stationers' Hall and at the National Archives. The SRO project has the support of the Register's custodians, the Worshipful Company of Stationers and Newspaper Makers.
- 2 The pilot SRO project is currently concluding. It has digitized Edward Arber's (1875–95) transcripts of the earliest volumes of the Register (1557–1640) and the later transcriptions of the volumes of 1640–1708, made by Eyre, Rivington, and Plomer in 1913–1914; estimated the cost of proofing and correction of the resulting transcription against the manuscript originals; and estimated the cost of transcribing later series from available sources (both manuscript and print). This initial project will produce enough data to demonstrate its usefulness, enabling future funding bids to complete the later material and create an online interface of the complete Stationers' Register. The eventual intention is to publish an indexed, searchable website containing full-text transcription of the complete Register. Such a website would make possible a linked interface between

this transcription and images of the original manuscript materials, should permissions for such a venture be granted. Neither the digitization of the manuscript books nor production of such a Web interface is within the scope of this pilot project.

- 3 The use and customization of the TEI recommendations is of particular interest in this project. The keying company chosen to transcribe the Register has a policy of charging per kilobyte of output produced, yet is content to work to any XML schema, providing that it is accompanied by encoding guidelines illustrated with examples, and that the source material's textual features required to be transcribed and marked up are easily recognizable by keyers. The money saved through creating a byte-reduced, highly abbreviated, and significantly constrained schema based on the TEI meant that the project could also produce transcriptions of the volumes of the Register edited by Eyre, Rivington, and Plomer. The focus of this paper (as of the project as it was originally conceived) is the transcription of Arber.
- 4 One of the joys of working in a small, hand-picked group is that skills and knowledge are complementary, and everyone's expertise informs the whole. It may be informative to outline the roles of the people involved in the SRO project. Giles Bergel and Ian Gadd collectively brought many years' academic research into the Stationers' Company and its records: their profound knowledge of the source material has guided all our work. Giles Bergel is also the project's Principal Investigator. James Cummings prepared the schema and its associated documentation, and was responsible for the project's technical development and transformation. Pip Willcox wrote the illustrated transcription guidelines on which the schema was founded, and managed the keying and editorial processes.

2. Source Material: Arber and the Stationers' Register

- 5 For the reasons outlined above, the SRO chose to create digital editions of Arber's nineteenth-century edition of the Stationers' Register. It has a reputation of being a complex and unwieldy set of volumes and would benefit from being accurately and fully searchable in digital form. As a more or less formulaic series of volumes (less in the early volumes, where the Register was used as a general accounts book of the Stationers' Company, and increasingly as the Register continues), it presents few encoding problems.
- 6 Recognizing this, Arber noted the usefulness of marking particular features consistently throughout the four Register volumes he edited so painstakingly. The tools at his disposal were page layout and choice of fonts, and he used these ("nineteenth-century XML encoding" as they were informally termed within the project) to indicate semantic categories of information. His editorial methods are detailed in the prefatory material, "On the Present Transcript" (Arber 1875–95, 1:27–30). It is his consistency of presentational markup, and the rendition of it by the keying company, that enables descriptive markup to be deduced and generated. Without this, our project could not have been contemplated in its present form.
- 7 A good illustration of this is his treatment of names. Arber estimates there are 40,000 real and fictitious names in the four volumes he edited, and he differentiates them in various ways (while noting the limitations of this work, due, for example, to members moving between livery companies). For example, "the names of all members, whether Freemen or Brethren, of the Stationers' Brotherhood or Company, so far as they could be ascertained,

[are set] in a special type technically known as 'Clarendon'" (Arber 1875–95, 1:27). "Names of persons generally, and in the Book Entries of authors real or fictitious, have been placed in Roman Small Capitals" (Arber 1875–95, 1:28). "Names of authors" appear in "Italic Capitals" (Arber 1875–95, 1:29).

- 8 Another factor that enabled this pilot project was the out-of-copyright status of Arber's and Eyre, Rivington, and Plomer's editions. Digital facsimiles of the Columbia University copy of Arber (1875–95)² and of the University of Toronto's copy of Eyre, Rivington, and Plomer (1913–14)³ are available online. Despite the project's name, it is worth remarking that we have created digital editions of these editions of the Register, rather than of the original manuscript fair copies. While the editors added much (particularly in Arber's case) to the manuscript, they are not complete transcriptions. Arber had permission only to transcribe entries relating to books, to the careers of individual printers, binders, publishers, and other members of the Company, and "dinner-bills 1557–1569, with some other similar items" (Arber 1875–95, 1:29).

3. Creating a Byte-Reduced TEI Schema (`tei_corset`)

- 9 Given the keying company's willingness to encode XML texts to any documented schema, with their outputs charged per kilobyte, the project realized that creating a byte-reduced schema would save the project money. A schema of this type would replace long and mostly human-readable names of elements, attributes, and their values with highly abbreviated forms. For example, the `<div>` element became `<d>`, the `@type` attribute became `@t`, and in most cases the allowed values for `@t` were tightly constrained. This means that what might be expanded as `<div type="entry">` (18 characters) was coded as `<d t="e">` (9 characters). The creation of such a schema is intended solely to reduce the number of characters used in the resulting edited transcription, as an intermediate step in the project's workflow. The decisions here sacrifice the laudable aims of human-readable XML for the sake of brevity. This is different from the TEI Tite customization intended for mass digitization which concentrated on providing a tightly-controlled and reduced selection of TEI elements for easier keying.
- 10 In addition, the number of elements allowed in the schema was greatly reduced. In total 34 elements were allowed in this internal schema:

Table 1: Elements in `tei_corset` customization

Module	Elements in <code>tei_corset</code>
textstructure	body div TEI
figures	cell row table
header	teiHeader
linking	ab seg
transcr	fw space
namesdates	forename surname

```

core          abbr add cb date foreign gap graphic head hi item
              label lb list name note num p pb q title unclear

```

- 11 The TEI customization was created using a TEI ODD file, and because it was significantly restricting the TEI we called it "tei_corset".⁴ The elements in question were referenced by inclusion rather than exclusion (TEI Consortium 2013, ch. 23). This ensured that were we to regenerate the produced schemas against the latest version of the TEI Guidelines, we would not be presented with new and unexpected elements. This portion of the TEI ODD file looks like this:

```

<schemaSpec ident="tei_corset" docLang="en" start="TEI" targetLang="en"
ns="http://www.tei-c.org/ns/corset/1.0">
  <!-- references to modules/elements to include -->
  <moduleRef key="tei"/>
  <moduleRef key="textstructure" include="body div TEI"/>
  <moduleRef key="figures" include="cell row table"/>
  <moduleRef key="header" include="teiHeader"/>
  <moduleRef key="linking" include="seg ab"/>
  <moduleRef key="transcr" include="fw space"/>
  <moduleRef key="namesdates" include="forename surname"/>
  <moduleRef key="core" include="abbr add cb date foreign gap graphic
head hi item label lb list name note num p pb q title unclear"/>
  <!-- ... -->
</schemaSpec>

```

4. Renaming Elements

- 12 In most cases the changes to the elements were fairly straightforward, such as renaming the element and @type attribute. For example, the <list> element was renamed to <ls> with the following <elementSpec>:

```

<elementSpec ident="list" mode="change">
  <altIdent>ls</altIdent>
  <equiv filter="corset-acdc.xml" mimeType="text/xml" name="list"/>
  <attList>
    <attDef ident="type" mode="change">
      <altIdent>t</altIdent>
    </attDef>
  </attList>
</elementSpec>

```

- 13 The <elementSpec> element uses the @ident attribute to identify the element being documented and the @mode attribute to explain what we are doing to it.⁵ In this case the element is undergoing a "change". One of the changes is to use <altIdent> to provide a new identifier for the element, that is, to rename it from <list> to <ls>. Using a

similar notation we change the attribute definition `<attDef>` for the `<list>` element's `@type` attribute to be `@t`.

- 14 There is an additional piece of literate programming documentation present here, and that is the use of the `<equiv>` element. This programming informs any ODD processor that a filter exists in a file called `corset-acdc.xml` (a file of Mime Type `text/xml`) called `list` which would revert to, or otherwise document, an equivalent notation. In this case, it is a named template in an XSLT file which reverses the changes in a given document that adheres to the modified schema. The XSLT template to rename this is trivial and might look like:

```
<xsl:template match="ls">
  <list>
    <xsl:apply-templates select="@*|node()" />
  </list>
</xsl:template>
```

- 15 In other cases the TEI ODD file not only renames the `@type` attribute but constrains entirely the values that are allowed in it. For example, in the `<name>` element (renamed `<n>`), the `@type` attribute (renamed `@t`) has a closed value list allowing only "per" (personal name), "pla" (place name), and "oth" (other name).

```
<elementSpec id="name" mode="change">
  <altId>n</altId>
  <equiv filter="corset-acdc.xml" mimeType="text/xml" name="name"/>
  <attList>
    <attDef id="type" mode="change">
      <altId>t</altId>
      <valList mode="replace" type="closed">
        <valItem id="per"><desc>personal name</desc></valItem>
        <valItem id="pla"><desc>place name</desc></valItem>
        <valItem id="oth"><desc>other name</desc></valItem>
      </valList>
    </attDef>
  </attList>
</elementSpec>
```

5. Non-Conformant Changes

- 16 Our `tei_corset` schema, like the `tei_tite` schema,⁶ is not in itself TEI-conformant: both make changes to the overall text structure which are contrary to the TEI Abstract Model as documented in the TEI Guidelines. The `tei_corset`, in addition to temporarily renaming the `<TEI>` element as `<file>`, changes the content model of the `<teiHeader>` element significantly. The element specification for the `<teiHeader>` element is as follows:

```

<elementSpec ident="teiHeader" mode="change">
  <altIdent>header</altIdent>
  <equiv filter="corset-acdc.xml" mimeType="text/xml" name="teiHeader"/>
  <content>
    <rng:ref name="title"/>
    <rng:zeroOrMore>
      <rng:ref name="model.pLike"/>
    </rng:zeroOrMore>
  </content>
  <attList>
    <attDef ident="type" mode="delete"/>
  </attList>
</elementSpec>

```

- 17 In other words, the `<teiHeader>` element is renamed `<header>` (as it was only used once per file, few characters would have been saved by abbreviating it further), and the locally defined `@type` attribute is deleted. Additionally, the content model is entirely replaced. Using embedded RELAX NG, this documents that a `<title>` element (later renamed `<t>`) is required, followed by zero or more members of the `model.pLike` class. This allowed the encoders to put a title for the file and any paragraphs noting problems or questions they had, but gave provision for no other metadata to be supplied in the header of the file. The reason for this is that the headers of each of the files were destined to be replaced with more detailed, TEI-conformant ones later in the project's workflow.
- 18 To note that these were not truly TEI-conformant files, the namespace for the whole file was changed to one created specifically for `tei_corset`. This namespace was replaced as part of the conversion back to TEI. There are other changes that are not TEI-conformant in the `tei_corset` customization. For example, both `<pb>` and `<gap>` (renamed as `<gp>`) are changed to allow textual content. While `<gap>` allows `<desc>` inside it, this was deliberately excluded from the `tei_corset` ODD. The gaps in Arber's edition are stated regularly as, for example, "[Four entries omitted]". Here as well, the regularity of Arber's notation made our work simple, this time in up-converting into XML markup; the notion of "up-conversion" here includes not only converting from one format to another, but also programmatically deriving the creation of additional intellectual value.

```

<elementSpec ident="gap" mode="change">
  <equiv filter="corset-acdc.xsl" mimeType="text/xsl" name="gap"/>
  <altIdent>gp</altIdent>
  <!-- allow gap to have text content -->
  <content>
    <rng:text/>
  </content>
  <attList>
    <attDef ident="agent" mode="delete"/>
    <attDef ident="hand" mode="delete"/>
    <attDef ident="reason" mode="delete"/>
  </attList>
</elementSpec>

```

6. Using the TEI Class System

- 19 As you can see in the example above, the local attributes `@agent`, `@hand`, and `@reason` are deleted from this element. While `<gap>` allows many more attributes in the TEI, these come from the TEI attribute class system. In `tei_corset`, these classes were either deleted or modified separately, so the elements which normally claim membership in them lost all the attributes those classes provide. For example, some of the classes deleted are:

```

<classSpec ident="att.ascribed" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.breaking" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.dataable.iso" mode="delete" module="tei" type="atts"/>
<
<classSpec ident="att.declaring" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.dimensions" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.duration" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.internetMedia" mode="delete" module="tei"
type="atts"/>
<classSpec ident="att.naming" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.placement" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.ranging" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.sourced" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.spanning" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.translatable" mode="delete" module="tei"
type="atts"/>
<classSpec ident="att.transcriptional" mode="delete" module="tei"
type="atts"/>
<classSpec ident="att.divLike" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.editLike" mode="delete" module="tei" type="atts"/>
<classSpec ident="att.segLike" mode="delete" module="tei" type="atts"/>

```


- 20 One of the classes which was modified significantly was the `att.global` class. Some of the elements (such as `@rendition`, `@xml:base`, and `@xml:space`) were deleted. Other elements were modified for brevity, such as `@xml:id` (renamed `@id`) and `@xml:lang` (renamed `@lg`).
- 21 The `@rend` attribute (renamed `@r`) was to be used particularly heavily, and was therefore given an extensive list of values. We allowed the attribute to appear in many different elements because the key to our up-conversion to semantic XML was a description of Arber's formatting. Table 2 shows allowable `@rend` values and their equivalent meanings.

Table 2: Allowable `@rend` attribute values in `tei_corset` customization

<code>@rend</code> (renamed <code>@r</code>) attribute value	Meaning
"ab"	rendered above the line
"al"	rendered aligned to the left
"ar"	rendered aligned to the right
"b"	rendered in bold
"bel"	rendered below the line
"bl"	rendered in blackletter font
"brl"	rendering is bracketed to the left
"brr"	rendering is bracketed to the right
"c"	rendered centred
"dc"	rendered as drop-cap or illuminated initial
"f"	rendered in a different font
"i"	rendered in italics
"l"	rendered on the left
"lrg"	rendering is of large size
"med"	rendering is of medium size
"n"	rendering returns to "normal"
"o"	other rendering
"r"	rendered on the right
"rm"	rendered in roman numerals
"s"	rendered in superscript

"sc"	rendered in small caps
"sig"	rendered as a signature
"sml"	rendered as smaller
"st"	rendered struck-through
"u"	rendered in underline
"xlg"	rendering is of extra-large size
"xxlg"	rendering is of extra-extra-large size

7. Fees, Roman Numerals, and Old-Style Money

- 22 One of the structural divisions of information that the keyers were marking up was a `<div type="entry">` (or, in its renamed form, `<div type="entry">`) which contained a single Register entry recording one or more works registered with the company. Inside such entries the fees paid are (almost) always aligned to the right margin and given in roman numerals. The keying company was asked to mark these numbers (the `<num>` element had been renamed `<nm>`) and to use a renamed `@rend` attribute (`@r`) to indicate its formatting. This meant that we could determine, fairly consistently, when they had marked registration fees paid to the Company. The keying company might mark a number as in the example below (in this case for timber for building work in 1569). The example below (`vli iijs iiijd`) is in fact an expense for building work, which we use here rather than a book title registration because it demonstrates the use of pounds, shillings, and pence. Most entries registering books at this time cost sixpence, but seeing a record with a larger amount of money (in pounds, shillings, and pence) is helpful to demonstrate the complexity of the following up-conversion which was needed in a few instances. This might have been keyed as:

```
<nm r="ar rm">vli iijs iiijd</nm>
```

- 23 In this case "ar" means it was rendered as aligned to the right, and "rm" that it was in roman numerals. This record mistakenly appears inside an 'entry' and so could falsely be assumed to be a book registration. Indeed, one of the reasons the data will need significant cleaning is that the keying company tended to use "entry" (the most common `<div>` `@type`) as a default, rather than for recording only the registration of works to be printed. Nonetheless, our task in converting the above number is not only to transform it back into the pure TEI element of `<num>` but to deduce additional information where possible. This means tokenizing the amounts into separate pounds, shillings, and pence, encoding the unit marker as superscript (Arber's edition did this so consistently that we did not ask the keying company to mark the feature, realizing we could convert it automatically later, saving more characters), and also convert the roman numerals to arabic numerals and provide a total in a common unit. In the XSLT stylesheet converting the keyed data back to pure TEI, breaking the string up based on whether something ended 'li' (pound), 's' (shilling), 'd' (pence), or 'ob' (halfpenny) with a space

following it was fairly easy to do consistently. These markers need to be removed from the amount being converted since some of them, such as 'li', could be interpreted as roman numerals. A slightly more complex XSLT function not only changes the roman numerals to arabic, but converts them to pence so that we can store the individual and aggregate amount. The resulting markup provides a great deal more information than the input above:

```
<seg type="fee" rend="roman-numerals aligned-right">
  <num type="totalPence" value="1240">
    <!--orig: vli iijs iiijd -->
    <num type="poundsAsPence" value="1200">v<hi rend="superscript">li</
hi></num>
    <num type="shillingsAsPence" value="36">iiij<hi rend="superscript">s</
hi></num>
    <num type="pence" value="4">iiij<hi rend="superscript">d</hi></num>
  </num>
</seg>
```

- 24 The fee is marked as a `<seg>` element (for consistency with other elements marking short portions of text in the converted data; the `<hi>` element could also have been used) with a `@type` attribute of "fee" (mistakenly in this case; it is actually a payment made by the Stationers and not a fee received) and the `@rend` classifications are expanded to be more human-readable. The original string passed to the conversion function is stored as an XML comment to aid proofreading. A set of nested `<num>` elements give the values for pounds, shillings, and pence, and an outer `<num>` element gives the "totalPence" value.⁷ The `<measure>` element could also have been used here for more detailed recording of information.⁸ Up-converting such information is always likely to be inexact when dealing with a large amount of messy data, but nonetheless can provide valuable information to be used for comparative financial study. Being able to find fees where the amount is uncharacteristically large would help to show the location of potential encoding errors such as this one so that they could be corrected.

8. Titles of Books

- 25 Given the complexity of the data (both the original manuscript Register and Arber's edition of it), it is unsurprising that other aspects of these transcriptions prove difficult to up-convert. For example, in entries registering one or more titles, we noted from the project's start that it would be useful for future exploitation of the data to mark these as titles. As noted above, Arber used an italic font for titles, and keyers were instructed to mark italicized passages as such. Arber's (conventional) typographic habits, for example of indicating foreign or formulaic words and other textual phenomena using an italic font, meant the up-conversion of these italicized passages into titles was not straightforward. Phrases such as "i.e.", "pp", "See pp.", "etc." are marked in italics and come with a wide variation of spellings and punctuation. Month names in Latin such as "januarij", "marcij" or "martij", and "septembris" are also recorded in italics, as are editorial phrases such as "no money payment", "no payment recorded", or cross-references such as "*vide proxima paginam*".

- 26 Checking the texts became a war of attrition against the mis-identification of such passages as titles. Each individual example was added to a variable in the conversion XSLT to be used as a look-up table for comparison. A further confounding feature was the keyers' overuse of "entry" as a `<div> @type`, so that a potential means of narrowing the number of italicized passages (excepting in Arber's notes, the only use of italicized fonts within entries is for book titles) was lost. Ideally every italicized passage, and every passage marked as a title, would be checked by hand, but resources in this phase of the project have not allowed this.
- 27 Our compromise is to leave what we consider an acceptable number of false positives in the transcriptions: more will be found during subsequent phases of the project, where work will encompass more than a proof-of-concept and initial data. What is unknown and will take significant proofreading is how many false negatives there are—how many times a title has not been marked either because the italics were not captured by the keying company, or because Arber himself accidentally omitted to italicize it.

9. Proofreading and Quality Analysis

- 28 "The actual process of all good correction is very slow" (Arber 1875–95, 1:27). Despite the high accuracy of the transcriptions the keying company provided, proofreading is desirable: early modern spelling is variable, and human errors can creep in; the digital facsimile is not always clear; the fonts are unusual to modern eyes, and some of them are hard to read because of their small print size.
- 29 Given the limited resources of the current phase, it was never the intention of the project to proof the entire transcription against the printed editions (even before Eyre, Rivington, and Plomer's volumes were included). Rather, we proofread a small sample of the transcription, and then used automated checks on much of the encoding. Subsequent phases of the SRO may proofread these transcriptions against the manuscript originals (or images of them).
- 30 Early in our checking, it became apparent that names needed particular attention. The keyers had been asked to mark up names and describe their formatting. We underestimated the complexity involved in identifying a name in early modern text for non-specialists: for example, places and people were confused; patron saints of churches were included as names; the `<name>` element did not always surround the entire name, sometimes including titles, sometimes not.
- 31 Because names are an important focus in research, we concentrated on proofreading the 40,000 that appear in Arber, correcting errors as we encountered them (rather than proofing a random sample of the whole transcription). This work is nearing completion, taking longer than anticipated as, like Arber, we now have "Leisure moments only ... available for its execution" (Arber 1875–95, 1:27).

10. Conclusions

- 32 There are many lessons learned from the pilot SRO project, but one of the main ones must be that careful customization of the TEI can deliver significant benefits, especially on projects with limited budgets, when dealing with a keying company that charges per

kilobyte of output. Although the data are not always entirely consistent, the information we are able to derive from what data we have is significantly better than no information.

- 33 Creating full-text transcriptions of the material covered in the pilot project directly from manuscript would not have been possible within the project's resources; the ability to work from a reliable, out-of-copyright transcription made our work possible. A quick estimate before the digitization suggested that we would save approximately 40 percent in file size from using the `tei_corset` schema. When combined with the up-conversion provided by deriving new markup and expanding the abbreviated values automatically, the saving in file size was closer to 60 percent. This saving meant we were able to include the Eyre, Rivington, and Plomer 1913–14 transcripts of the Register (1640–1708) as part of the project. However, some of that saved money may have to be balanced against the need for additional proofreading.
- 34 It seems fitting, given how reliant the SRO project has been on Arber's editions, to echo him as we look forward to the publication of these transcriptions. Arber's work was solitary. By contrast, ours has been collaborative from the outset. We hope to extend that collegiality further, and that the academic community will eventually be able help improve our transcriptions: "One final request must here be made as this Volume passes into the hands of the Reader, viz: That its contents may be *at once* subjected to the severest possible scrutiny; so that any hitherto unapparent errors may be corrected" (Arber 1875–95, 1:30).

BIBLIOGRAPHY

Arber, Edward, ed. 1875–1895. *A Transcript of the Registers of the Company of Stationers of London 1554–1640 AD*, 5 vols. London & Birmingham: private printing.

Eyre, George Edward Briscoe, Charles Robert Rivington, and Henry Robert Plomer, eds. 1913–14. *A Transcript of the Registers of the Worshipful Company of Stationers from 1640–1708*, 3 vols. London: private printing.

Gadd, Ian Anders. 2004. "Were Books Different? Locating the Stationers' Company in Civil War London, 1640–1645." In *Institutional Culture in Early Modern Europe*, edited by Anne Goldgar and Robert Frost. Leiden: Brill.

TEI Consortium. 2013. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.5.0. Last updated July 26. N.p.: TEI Consortium. <http://www.tei-c.org/Vault/P5/2.5.0/doc/tei-p5-doc/en/html/index.html>.

NOTES

1. The principal investigator is Dr. Giles Bergel (University of Oxford), and the project has benefited from working closely with Professor Ian Gadd (Bath Spa University). The authors are indebted to both these colleagues for their generosity in sharing their

research and extensive knowledge of the Stationers' records and their context, particularly in the preparation of the posters and papers we have presented at several conferences. The project has also benefited from the support of the Worshipful Company of Stationers and Newspaper Makers, in our ongoing work, and in granting permission to reproduce images of the Register in conference papers and posters. We remain grateful to the Lyell Research Fund, which made our work possible.

2. *A transcript of the registers of the company of stationers of London 1554-1640, A.D.*, ed. Edward Arber, vols. 1-5 (New York: Columbia University Libraries, 2007), accessed January 7, 2014, <http://clio.cul.columbia.edu:7018/vwebv/holdingsInfo?bibId=6177070>.

3. *A Transcript of the Registers of the Worshipful Company of Stationers, from 1640-1708, A.D.* (Internet Archive), accessed January 7, 2014, <http://archive.org/details/1913transcriptof03statuoft>.

4. The original `tei_corset` ODD file and `corset-acdc.xsl` stylesheet are available at https://github.com/jamescummings/conlucvies/tree/master/tei_corset.

5. TEI Consortium 2013, appendix C, `<elementSpec>`, <http://www.tei-c.org/Vault/P5/2.5.0/doc/tei-p5-doc/en/html/ref-elementSpec.html>.

6. Perry Trolard for the TEI Consortium, "TEI Tite: A Recommendation for Off-site Text Encoding," Version 1.1, September 2011, http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html.

7. TEI Consortium 2013, Appendix C, `<num>`, <http://www.tei-c.org/Vault/P5/2.5.0/doc/tei-p5-doc/en/html/ref-num.html>.

8. TEI Consortium 2013, Appendix C, `<measure>`, <http://www.tei-c.org/Vault/P5/2.5.0/doc/tei-p5-doc/en/html/ref-measure.html>.

ABSTRACTS

This article examines the process of customization, digitization, up-conversion, and proofreading that took place as part of the Stationers' Register Online (SRO) project at the University of Oxford. This pilot SRO project received institutional funding to digitize the first four volumes of the book-entry Register of the Stationers' Company. This is an important source for book history in Britain as the Stationers' Register was the primary means through which ownership of texts was asserted, disputed, regulated, and monitored from 1577 until 1924. The article concentrates on the technical aspects of creating a byte-reduced TEI schema (`tei_corset`) which, because of the arrangement with the keying company, saved a significant amount of money. It looks at the creation of that schema, the up-conversion of the resulting XML, and the difficulties of proofreading such a large quantity of material.

INDEX

Keywords: customization, ODD, conversion, digitization, book history