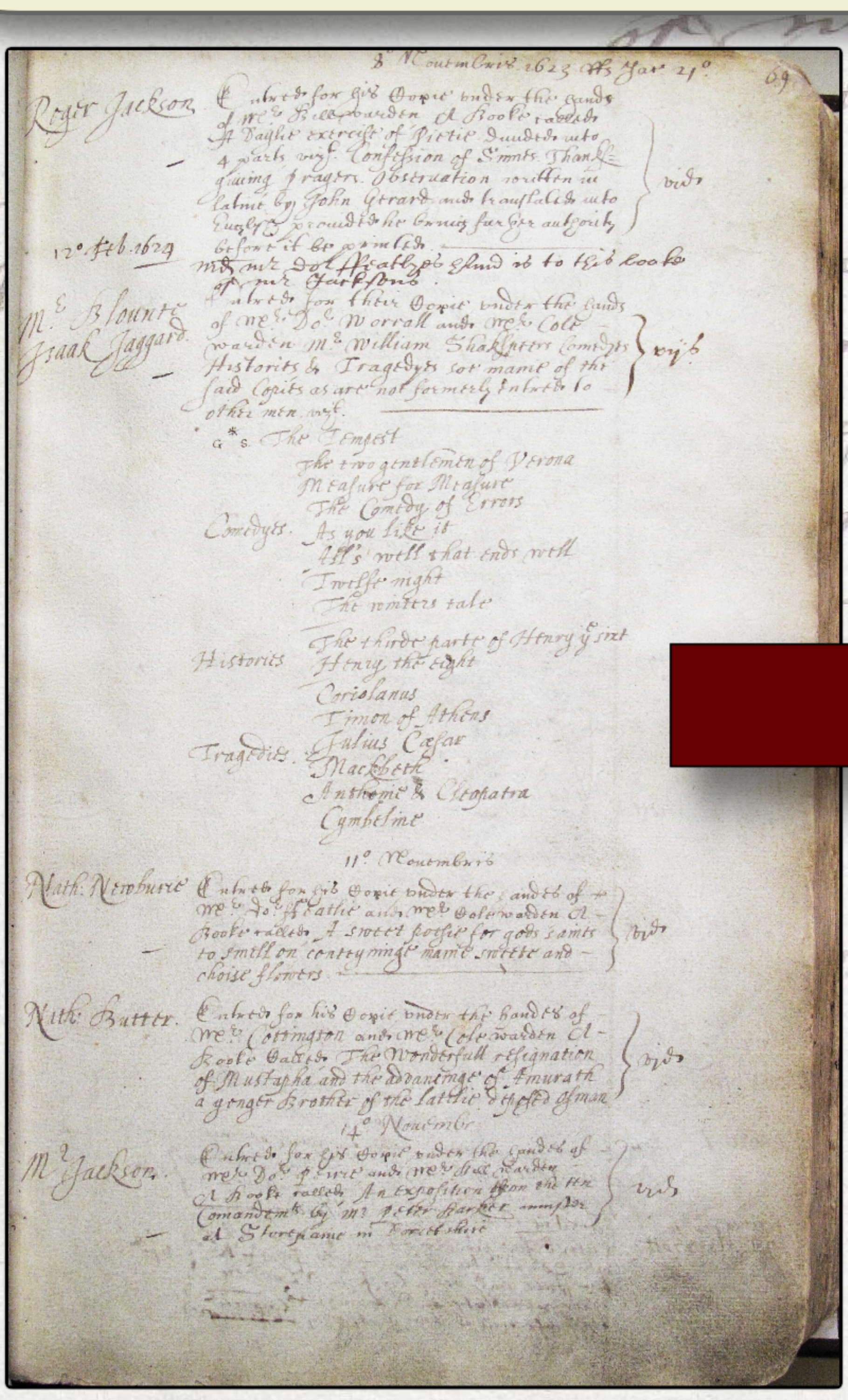
Digitizing the Stationers' Register

This poster examines the technical background to the Stationers' Register Online (SRO) project based at the University of Oxford. The pilot SRO project received institutional funding from the Lyell Research Fund to transcribe Arber's and Eyre and Rivington's editions of the Register of the Stationers' Company, (1554 to 1640, and 1640 to 1708) into XML based on the Guidelines of the Text Encoding Initiative (TEI).

As part of the preparation for the digitization of the earliest volumes of the Register by a keying company, the project created a byte-reduced encoding schema for concise digitization without loss of intellectual content. The motivating factor for this is that the keying company charges per kilobyte of output. While XML output was desired, there were significant savings to be had from reducing file size: around a 66% reduction. The transformation stylesheets to convert the transcriptions to "pure" TEI P5 XML from the reduced version created for the project also included automatic recognition and transformation of dates, and fees in Roman numerals. After this significant proofreading was undertaken by the project to correct errors introduced in the data capture and conversion.

The SRO project used an approach based on the digitization of Arber's print edition, and the presentational and editorial markup it contains, as a successful route for getting useful TEI XML for further interrogation and exploitation.



Arber's Edition

[G. Cole. J. Bill. } G. Swinhowe. 8—14 November 1623] 69

Regni Regis Jacobi 2j°.

8° Nouembris 1623

Roger Jackson. Entred for his Copie vnder the hands of Master **Bill** warden, A Booke called *A Daylie exercise of Pietie Diuided into 4 parts, viz^t. Confession of Sinnes. Thanksgiuing*[.] *Prayers. Obseruation* written in Latine by JOHN GERARD, and translated into English PROUIDED he bring further authority before it be printed . . vj^d

12° Februarij. 1624 [i. e. 1625.] MEMORANDUM Doctor FFEATLYES hand is to this booke of master **Jacksons**

Master Blounte Isaak Jaggard. Entred for their Copie vnder the hands of Master Doctor WORRALL and Master **Cole** warden Master WILLIAM SHAKSPEERS *Comedyes Histories, and Tragedyes* soe manie of the said Copies as are not formerly entred to other men. viz^t. vij^s

COMEDYES. *The Tempest* / *The two gentlemen of Verona* / *Measure for Measure* / *The Comedy of Errors* / *As you like it* / *All's well that ends well* / *Twelfe night* / *The winters tale*

HISTORIES *The thirde parte of HENRY ye SIXT* / *HENRY the EIGHT*

TRAGEDIES. *CORIOLANUS* / *TIMON of Athens* / *JULIUS CÆSAR* / *MACKBETH* / *ANTHONIE and CLEOPATRA* / *CYMBELINE*

11° Nouembris

Nathaniel Newburie. Entred for his Copie vnder the handes of Master Doctor FFEATLIE and Master **Cole** warden, A Booke called, *A sweet poesie for GODs saints to smell on, conteyninge manie sweete and choise flowers.* vj^d

Nathaniel Butter. Entred for his Copie vnder the handes of Master COTTINGTON and Master **Cole** warden, A Booke Called, *The Wonderfull resignation of MUSTAPHA, and the aduancinge of AMURATH a yo[u]nger Brother of the latelie deposed OSMAN* . . . . . vj^d

14° Nouembris

Master Jackson. Entred for his Copie vnder the handes of Master Doctor PEIRCE, and Master **Bill** warden A Booke called, *An exposition vpon the ten Comandements* by master PETER BARKER minister at Storepaine in Dorcetshire . . . . vj^d

IV. 107

The Stationers' Register is arguably the most important primary source for the study of the history of the book in Britain other than the books themselves. The Register was the primary means through which ownership of texts was asserted, disputed, regulated and monitored from c.1557 until 1924, and survives intact in two series now held in Stationers' Hall and at the National Archives.

Arber's volume presents information he gleaned from his reading of the register in font and stylistic changes. For example, a name in bold is a known member of the Stationers' Company whereas small-caps is used to indicate a non-Stationer. Arber is very consistent in his presentation of dates and regularization of the manuscript layout. SRO can use this presentational markup to derive semantic markup and useful metadata.

Create Custom TEI Schema

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

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

A TEI ODD customisation file enabled us to rename elements like <list> to be <ls> or the 'type' attribute to be merely 't'. As the keying company was charging per kilobyte of output the SRO project saved 66% by using a custom byte-reduced schema it created called 'tei_corset'. This ODD file also includes project specific documentation and encoding guidelines.

Transcribe To Schema

```xml
<d t="e">
   <hd r="l"><n r="b">Master <sn>Blounte</sn></n>
   <n r="b"><fn>Isaak</fn> <sn>Jaggard</sn></n>.</hd>
   <p>Entred for their Copie vnder the hands of Master Doctor <n r="sc"><sn>Worrall</sn></n>
      and <n r="b">Master <sn>Cole</sn></n> warden
      <n r="sc">Master <fn>William</fn> <sn>Shakspeers</sn></n>
 <h r="i">Comedyes Histories, and Tragedyes</h> soe manie of the said Copies as are not
formerly entred to other men. <h r="i">viz<h r="s">t</h></h>.
      <nm r="rm ar">vijs</nm></p>
   <p r="al"><h r="sc">Comedyes</h></p>
   <ls>
      <it><h r="i">The Tempest</h></it>
      <it><h r="i">The two gentlemen of Verona</h></it>
      <it><h r="i">Measure for Measure</h></it>
      <it><h r="i">The Comedy of Errors</h></it>
      <it><h r="i">As you like it</h></it>
      <it><h r="i">All's well that ends well</h></it>
      <it><h r="i">Twelfe night</h></it>
      <it><h r="i">The winters tale</h></it>
   </ls>
   <p r="al"><h r="sc">Histories</h></p>
    <ls>
     <it><h r="i">The thirde parte of</h> <n r="sc i"><sn>Henry</sn></n>
        <h r="i">ye</h> <h r="sc i">Sixt</h></it>
     <it><n r="sc i"><sn>Henry</sn></n> <h r="i">the</h> <h r="sc"><h r="i">Eight</h></h></it>
    </ls>
   <p r="al"><n r="sc"><sn>Tragedies</sn></n>.</p>
   <ls>
```

The customized 'tei_corset' format used XML as a markup language but with a highly abbreviated vocabulary. Although <d t="e"> is legible as plain text it is not as 'human-readable' as <div type="entry">. Similarly <nm r="rm ar">vijs</nm> is not very easily processed for financial statistical analysis but converting and storing all fees to a standard pence value in metadata will enable this.

Convert With XSLT

```xml
<!-- names -->
<xsl:template match="n[not(@t) or @t='per']" >
   <persName>
      <xsl:apply-templates select="@*|node()|comment()|processing-instruction()|text()"/>
   </persName>
</xsl:template>

<xsl:template match="n[@t='pla']">
   <placeName>
      <xsl:apply-templates select="@*|node()|comment()|processing-instruction()|text()"/>
   </placeName>
</xsl:template>

<xsl:template match="n" priority="-1">
   <name>
      <xsl:apply-templates select="@*|node()|comment()|processing-instruction()|text()"/>
   </name>
</xsl:template>

<!-- process numbers  that are inside entries and roman-->

 <xsl:template match="d[@t='e']//nm[contains(@r, 'rm')][contains(@r, 'ar')]" priority="1">
   <seg type="fee">
      <xsl:apply-templates select="@*"/>
      <xsl:copy-of select="jc:OldMoneyToTEI(.)"/>
   </seg>
 </xsl:template>
```

In addition to reverting all the abbreviated XML back to full TEI P5 XML, the XSLT conversion by the project also converted Roman numeral fees to pence. So the <nm> element above comes out as <num type="shillingsAsPence" value="84">vij<hi rend="superscript">s</hi></num>. Similarly, where possible dates have been tidied up and also stored as an attribute in a standard YYYY-MM-DD format.

Proofread Resulting TEI XML

```xml
<div type="entry">
 <head rend="left">
   <persName rend="bold">Master <surname>Blounte</surname></persName>
   <persName rend="bold"><forename>Isaak</forename> <surname>Jaggard</surname></persName>.
 </head>
 <p>Entred for their Copie vnder the hands of Master Doctor <persName rend="smallcaps">
    <surname>Worrall</surname></persName> and
    <persName rend="bold">Master <surname>Cole</surname></persName> warden
    <persName rend="smallcaps">Master <forename>William</forename> <surname>Shakspeers</surname>
   </persName>
   <hi rend="italics">Comedyes Histories, and Tragedyes</hi> soe manie of the said Copies as are
    not formerly entred to other men. <hi rend="italics">viz<hi rend="superscript">t</hi></hi>.
    <seg type="fee" rend="roman-numerals aligned-right">
       <num type="totalPence" value="84">
         <num type="shillingsAsPence" value="84">vij<hi rend="superscript">s</hi></num>
       </num>
    </seg>
 </p>
 <p rend="aligned-left">
   <hi rend="smallcaps">Comedyes</hi>
 </p>
 <list>
   <item><hi rend="italics">The Tempest</hi></item>
   <item><hi rend="italics">The two gentlemen of Verona</hi></item>
   <item><hi rend="italics">Measure for Measure</hi></item>
   <item><hi rend="italics">The Comedy of Errors</hi></item>
   <item><hi rend="italics">As you like it</hi></item>
   <item><hi rend="italics">All's well that ends well</hi></item>
   <item><hi rend="italics">Twelfe night</hi></item>
```

The resulting TEI XML will be invaluable for research into the history of copyright and publishing in Britain. Statistics can be derived about the frequency of registration; those registering works; the frequency of certain terms and conditions; and the value of registrations over a period of time. Future development may link the data to bibliographical records and to biographical resources for the study of the book trade.

The SRO project team consists of Ian Gadd (Bath Spa University), Giles Bergel, James Cummings and Pip Willcox (University of Oxford). For more information on the project, email stationersregisteronline@gmail.com. Images of the Stationers' Register are reproduced with the kind permission of the Worshipful Company of Stationers and Newspaper Makers. Image of the Arber transcript is taken from Edward Arber, ed., *A Transcript of the Registers of the Company of Stationers 1554-1640 AD.*, 5 vols (London & Birmingham, 1875-94), vol. 4, p. 107. The project has also transcribed G. E. Briscoe Eyre and C. R. Rivington, eds., *A Transcript of the Registers of the Company of Stationers from 1640-1708 A.D.*, 3 vols (London, 1913-14).

BATH SPA UNIVERSITY

Bodleian Libraries UNIVERSITY OF OXFORD

UNIVERSITY OF OXFORD