Using Stand-off XML Markup to Record Scholarly Differences of Opinion About Typesetting

by Gabriel Egan

Citation

Abstract

Not every set of questions we want to answer about literary/historical texts is best addressed by comprehensive transcription of all the features of the documentary witnesses. There is much that can be done in XML without undertaking a comprehensive encoding, and this paper will illustrate one such roll-your-own XML application for answering questions about the compositors who typeset the first editions of Shakespeare. Forty-six years ago Fredson Bowers foresaw computers transforming this field: "It will be a blessed day in the future when one can press a button and give such a lordly command as 'List for me every time compositor B follows his copy in spelling win as win or winne, every time he changes a copy spelling win to winne, or winne to win, and distinguish in each case what he does in setting prose and setting verse'". We've been rather slow at addressing this problem, and existing transcriptions of the early editions are distinctly unhelpful here. The paper will show how far a lone scholar can get towards a tool that will answer Bowers's questions using freely available materials and just a little knowledge of XML.

Using Stand-off XML Markup to Record Scholarly Differences of Opinion About Typesetting

by Gabriel Egan

The great bibliographer Fredson Bowers was an incorrigible optimist. In 1966 he had a vision of his field being transformed:

> I have some hopes that electronic computers can be put to work to digest and to analyse much information that at present we do not have. It will be a blessed day in the future when one can press a button and give such a lordly command as 'List for me every time compositor B follows his copy in spelling *win* as *win* or *winne*, every time he changes a copy spelling *win* to *winne*, or *winne* to *win*, and distinguish in each case what he does in setting prose and setting verse. Then give me all the occurrences of *win* and *winne* in texts that he set from manuscript'.[1]

Bowers was referring to the act of typesetting in which a compositor read the work he was supposed to set in type (an existing book or manuscript) and picked individual letters and punctuation from a typecase and placed them together word by word and line by line to make a block of printable type. Quite often the job of typesetting one book would be shared by several compositors and we can tell where each one started and finished his stint because they varied in their habits of spelling and spacing of type. For almost all books we do not know the compositors' personal names so they are identified as compositor A, compositor B, and so on.

Bowers's hopes have not yet been realised. The problem is tougher than he anticipated because where two or more scholars have tried to distinguish the compositorial stints in one book they have come to different conclusions about the numbers of compositors at work and where each one's stints start and finish. If we are to computerise the scholarly knowledge, we will have to record it as a set of hypotheses so that our questions take the form "supposing that Paul Werstine is right about his stints, where does compositor B spell *win* as *winne*?" and "now show me the same supposing

that Gary Taylor is right". That is, we have to computerise the scholarly differences of opinion.

The currently most popular way to computerise knowledge about a written text is to take an electronic version of the raw words and punctuation and to surround various parts of it with tags that conform to a standard known as Extensible Markup Language (XML). Just what features one records and what names one gives the tags are up to the user, but a typical model for marking up a play would be:

```
<play>
<act n="1">
<scene n="1">
 <line n="1">Bar. WHose there?</line>
<line n="2">Fran. Nay answere me. Stand and vnfolde your selfe.</line>
<line n="3">Bar. Long liue the King,</line>
 . . .
  <line n="156">Where we shall find him most conuenient. Exeunt</line>
 </scene>
</act>
. . .
 <act n="5">
. . .
 </act>
</play>
```

Each component part of the play is marked by a pair of tags, the opening one naming which kind of component it is, such as play, act, scene or line, and the closing one repeating the name but prefixed by a back-slash meaning "end of" play, act, scene or line. An important point is the Russian-doll (or Chinese-box) principle: the lines are nested inside scenes, which are nested inside acts, which are nested inside the outermost box called "play". This nesting is demanded in XML – no line may cross a scene boundary, no scene may cross an act boundary – because XML treats every text as what is called an Ordered Hierarchy of Content Objects. This means that in XML all novels have to consist of chapters that consist of paragraphs, and poems have to consist of lines that are made of words. For that reason XML has trouble

with the works of writers such as Laurence Stern, who had his printer put marbled endpaper in the middle of his novel *Tristram Shandy*,[2] or E. E Cummings who frequently broke words across line boundaries, as here at the beginning of his poem "exit a kind of unkindness exit":

exit a kind of unkindness exit

little
mr Big
notbusy
Busi
ness notman[3]

We might suppose that such violations of the Russian-doll principle are rare outside of literary conceit, but Murray McGillivray recently pointed out a typical example from everyday email:

Subject: Your parcel
From: Helen Black heblack@ucalgary.ca
Date: Fri. October 2, 2009 3:28 p.m.
To: Murray McGillivray <mmcgilli@ucalgary.ca>
Message: went off in the courier this afternoon, HEB[4]

The sentence "Your parcel went off in the courier this afternoon" is split between two containers: the 'Subject' and the 'Message'.The features of a book that a bibliographer is interested in, such as pages, formes and gatherings, cut across the features usually marked up in XML such acts, scenes, speeches and lines. A speech may easily cross a page boundary and a scene a forme boundary. There are well established means to reconcile incompatible hierarchies of interest within one XML document, but recording scholarly opinions about compositorial stints is a particularly tough case. Suppose that Werstine thinks that compositor A set the first line of the second quarto of Shakespeare's *Hamlet* (1604-5) and compositor B the rest of the play. We might mark this up thus:

```
<werstine-stint comp="A">
 <line n="1">Bar. WHose there?</line>
</werstine-stint>
```

```
<werstine-stint comp="B">
<line n="2">Fran. Nay answere me. Stand and vnfolde your
selfe.</line>
<line n="3">Bar. Long liue the King,</line>
. . .
</werstine-stint>
```

This is acceptable (technically, "well-formed") XML: each line is wholly contained within one of the two stints as determined by Werstine. Now let us suppose that Taylor thinks that compositor A set the first two lines of *Hamlet* and compositor B the rest of the play. We would mark this up thus:

```
<taylor-stint comp="A">
<line n="1">Bar. WHose there?</line>
<line n="2">Fran. Nay answere me. Stand and vnfolde your
selfe.</line>
</taylor-stint>
<taylor-stint comp="B">
<line n="3">Bar. Long liue the King,</line>
. . .
</taylor-stint>
```

This also is well-formed XML: each line is wholly contained within one of the two stints as determined by Taylor. Each of these two hierarchies exists perfectly well within its own XML document, but a problem occurs if we try to make them co-exist in a single document:

```
<werstine-stint comp="A">
<taylor-stint comp="A">
<line n="1">Bar. WHose there?</line>
 </werstine-stint>
<werstine-stint comp="B">
 <line n="2">Fran. Nay answere me. Stand and vnfolde your
selfe.</line>
</taylor-stint>
<taylor-stint comp="B">
<line n="3">Bar. Long liue the King,</line>
. . .
```

</taylor-stint>
</werstine-stint>

This is no longer well-formed XML because we have broken the Russian-doll/Chinese-box principle, or as they say in XML circles we have created overlapping hierarchies. It appears that we cannot make a single representation of Q2 *Hamlet* containing at once Werstine's and Taylor's views on its typesetting.

We are forced, then, to have one document for Werstine's view and one for Taylor's. But we do not want two complete copies of the play itself, not least because if we find a transcription error in the electronic text we do not want to have to correct it in two places. We should instead store in one document the base text, with the markup that everyone agrees upon, and keep the scholars' competing views of it somewhere else. This approach is called stand-off markup. Here are snippets from the five documents needed for Q2 *Hamlet*:

<line n="TLN-1">Bar. WHose there?**</line>**
<line n="TLN-2">Fran. Nay answere me. Stand and vnfolde your selfe.**</line>**
<line n="TLN-3">Bar. Long liue the King,**</line>**
(basetext.xml)

<xi:include href="basetext.xml" pointer ="TLN-1">
(werstine-on-comp-A.xml)

<xi:include href="basetext.xml" pointer ="TLN-2">
<xi:include href="basetext.xml" pointer ="TLN-3">
(werstine-on-comp-B.xml)

<xi:include href="basetext.xml" pointer ="TLN-1">
<xi:include href="basetext.xml" pointer ="TLN-2">
(taylor-on-comp-A.xml)

<xi:include href="basetext.xml" pointer ="TLN-3">
(taylor-on-comp-B.xml)

The base text contains only the uncontroversial line information for the first three lines. The document giving Werstine's view on compositor A's setting

of those three lines simply picks out the first line, and the document giving his view of compositor B's setting of those three lines picks out the second two. The document giving Taylor's view of compositor A's setting picks out the first two lines, and the document giving his view of compositor B's setting picks out just the third one. In this form, the documents holding the scholars' views can be interrogated by off-the-shelf software running the system called XQuery and during processing these XInclude statements are replaced with the content they identify, thus:

BEFORE PROCESSING
<xi:include href="basetext.xml" pointer ="TLN-2">
<xi:include href="basetext.xml" pointer ="TLN-3">
(werstine-on-comp-B.xml)

AFTER PROCESSING
<line n="TLN-2">Fran. Nay answere me. Stand and vnfolde your selfe.**</line>**
<line n="TLN-3">Bar. Long liue the King,**</line>**
(werstine-on-comp-B.xml)

By running our XQuery question against the document "werstine-on-cop-B.xml" we are running it against just the parts of the play that Werstine thinks compositor B set in type. It would be tedious to write an XInclude statement for each line that Werstine thinks compositor B set, but we do not have to specify individual lines: the procedure works just as well for whole pages and even gatherings, so long as we have identified those elements in the base text.

Let us return to what Bowers wanted to be able to ask a computer, so we can plan just what we have to mark up in the base text:

List for me every time compositor B follows his copy in spelling *win* as *win* or *winne*, every time he changes a copy spelling *win* to *winne*, or *winne* to *win*, and distinguish in each case what he does in setting prose and setting verse. Then give me all the occurrences of *win* and *winne* in texts that he set from manuscript.

This kind of enquiry requires that we record what the compositor was looking at as his copy text when setting type: not only whether it was an existing printed book or a manuscript, but also exactly what its readings

were in the case of every word. This last requirement is a tall order since it means marking up another document, the copy text, and providing a word-level link between every word in the copy text and every word in the book made from it, so that the departures from copy text spelling can be determined. In the case of Shakespeare, Bowers's first lordly command must refer only to printed copy because there survive no manuscripts used to set the plays and from which we might recover the copy spellings. We can be sure an edition is a reprint of a preceding edition only in cases where the reprinting is so faithful that it repeats the errors in its copy, and luckily in these cases the two editions will be so alike that a computer can identify for us which word in the earlier edition matches which word in the later. However, by definition such a reprint would not be substantive and hence of lesser interest than editions printed directly from manuscripts. The second of Bowers's lordly commands refers to editions set from manuscripts, but he is careful not to ask for copy spellings since in Shakespeare's case these are largely unknown: all we know is that the copy was a lost manuscript.[5]

It may be that Werstine and Taylor have differing opinions about the nature of the printer's copy for a book, or certain pages or just certain lines of it. The place to store that information is not the agreed base text but the document recording a scholarly opinion about it, like this:

```
BEFORE PROCESSING
<xi:include   href="basetext.xml"   pointer   ="B1r"
copy="ms">
<xi:include   href="basetext.xml"   pointer   ="B1v"
copy="print">
```
(werstine-on-comp-B.xml)

Unfortunately this does not work: the resulting pages do not pick up the copy attribute. We can specify the copy at the beginning of the document "werstine-on-comp-B.xml" so that it covers all of what compositor B is supposed by Werstine to have set, but in fact Werstine might reasonably suppose that compositor B used different kinds of copy in different parts of the book. We can specify the copy within the pages or lines of the base text so that it applies equally to Werstine's and Taylor's analyses, but in fact Werstine and Taylor might reasonably disagree on this point. Chalk up one failure for this method.

Bowers's reference to different spellings in prose and verse arises from the

compositors' art of justification. When setting prose a compositor would insert additional small spaces between words to push the last word to the end of the line and so produce a smooth right edge to the page, whereas when setting verse he would use regular spaces between words and fill the end of the line with larger ones to give the page a jagged right edge. When the adjustment of the small spaces between words failed to fully justify a line of prose the compositor was free to alter the spellings of the words to get a tight fit, whereas in verse the expanse of space at the end of the line made this exigent unnecessary. Thus in prose – or indeed long verse lines that fill the measure – we cannot assume that the compositor's spelling choices reflect his personal preferences, since he might have resorted to them only to justify the line. In studying compositors' habits, then, it is useful to have a record of whether each line is full. This information is not controversial and we can simply add it as a second attribute of each line in the base text, thus:

```
<line n="TLN-1" length="not-full">Bar. WHose there?</line>
<line n="TLN-2" length="full">Fran. Nay answere me. Stand and vnfolde your selfe.</line>
<line n="TLN-3" length="not-full">Bar. Long liue the King,</line>
```
(basetext.xml)

Writing in 1966, Bowers confined himself to compositors' spelling preferences and did not consider psycho-mechanical habits – such as failure to insert spaces after commas in short lines (where justification cannot be the cause) – that T. H. Howard-Hill, McDonald P. Jackson and Gary Taylor later used to distinguish compositors.[6] Although it is not demonstrated here, such tests can be incorporated into the present methodology by adding to the base text special characters representing, for example, terminally spaced commas.

We have now provided enough information to ask the computer pertinent questions and may illustrate the method with a real-world example. The earliest surviving (although not necessarily the first) edition of Shakespeare's *Love's Labour's Lost* is a quarto of 1598. George R. Price identified three compositors at work in this edition, with the following division of labour by pages set:

Comp I A2r, A2v, A3r, A4r, A4v, B2r, B3v, Clr, C1v, C2r, C2v, F2r, G4r, G4v, H3r, H3v, H4v

Comp II A3v, B1r, B1v, B2v, B3r, C3r, D1v, D2r, D2v, D3v, D4v, E4r, E4v, F1r, F1v, F2v, G1r, G1v, G2r, G3r, G3v, H1r, H2v, H4r, I1r, I2r, I3r, I3v, I4r, I4v, Klr, K1v

Comp III B4r, B4v, C3v, C4r, C4v, D1r, D3r, D4r, E1r, E1v, E2r, E2v, E3r, E3v, F3r, F3v, F4r, F4v, G2v, H1v, H2r, I1v, I2v, K2r, K2v[7]

Paul Werstine also found three compositors at work in this edition, but with a quite different division of labour:

Comp R B1r

Comp S A2r, A2v, A3r, A3v, F1r, F1v, F2r

Comp T A1r, A1v (blank), A4r, A4v, B1v, B2r, B2v, B3r, B3v, B4r, B4v, C1r, C1v, C2r, C2v, C3r, C3v, C4r, C4v, D1r, D1v, D2r, D2v, D3r, D3v, D4r, D4v, E1r, E1v, E2r, E2v, E3r, E3v, E4r, E4v, F2v, F3r, F3v, F4r, F4v, G1r, G1v, G2r, G2v, G3r, G3v, G4r, G4v, H1r, H1v, H2r, H2v, H3r, H3v, H4r, H4v, I1r, I1v, I2r, I2v, I3r, I3v, I4r, I4v, K1r, K1v, K2r, K2v[8]

The first step is to find an electronic text of this early edition, and happily Michael Best's Internet Shakespeare Editions website has a suitable one. We could get it already marked up with the tags that Best uses, but we want just the raw words. Scraping the play's words off the screen (using CTRL-a, CTRL-c on a Microsoft Windows system) gives raw text that has unwanted extra matter at the top and the bottom of the document, which must manually be deleted. There are also unwanted line numbers in it and the occasional blank line to delete. A small program written in the language Perl can chop those out and wrap the tags around each line to make it a *line* element with the attributes *linenumber* and *length*, the latter set by default to 'not-full'. All that remains to make this a usable base text is manually to add tags marking the book's sheets and pages and to set the length attribute to 'full' where necessary. This we can do in an XML editor such as Oxygen using a facsimile of the edition as a crib.[9] The total manual editing time for this one play was around two hours.

Oxygen has an XQuery processor built in, so directly inside this editor we may interrogate the documents that represent Price's and Werstine's beliefs about which compositor set which part and thus we can give Bowers's lordly commands. Here is an XQuery asking for the full-length lines Werstine thinks were set by compositor S, together with the result it produces:

QUERY: doc("werstine's-comp-S.xml")//line[@length="full"]

```
<?xml version="1.0" encoding="UTF-8"?>
<line      xmlns:xi="http://www.w3.org/2001/XInclude"
linenumber="_13" length="full">LET Fame, that all hunt after in
their lyues, </line>
<line      xmlns:xi="http://www.w3.org/2001/XInclude"
linenumber="_14" length="full">Liue registred vpon our brazen
Tombes, </line>
<line      xmlns:xi="http://www.w3.org/2001/XInclude"
linenumber="_68" length="full">Ferd. Why that to know which
else we should not know. </line>
. . .
<line      xmlns:xi="http://www.w3.org/2001/XInclude"
linenumber="_1569" length="full">Duma. Darke needes no
Candles now, for darke is light. </line>
<line      xmlns:xi="http://www.w3.org/2001/XInclude"
linenumber="_1584" length="full">King. Then leaue this chat,
and good Berowne now proue </line>
```

The display of the line numbers gives us a useful check to ensure that the query is doing what we think it is doing. Then, the XQuery can be tweaked to get just the words in the lines without the surrounding XML tags:

QUERY: doc("werstine's-comp-S.xml")//line[@length="full"]/text()

```
<?xml version="1.0" encoding="UTF-8"?>LET Fame, that all
hunt after in their lyues, Liue registred vpon our brazen Tombes,
Ferd. Why that to know which else we should not know. Ber.
Things hid &amp; bard (you meane) from cammon sense. Lon. He
weedes the corne, &amp; still lets grow the weeding. Ber. The
Spring is neare when greene geese are a bree- Bero. Well, say I
am, why should proude Sommer boast, Bero. No my good Lord, I
```

haue sworne to stay with you. Fer. How well this yeelding rescewes thee from shame. Ber. Item, That no woman shall come within a myle of Ber. Lets see the penaltie. On payne of loosing her tung. Item, Yf any man be seene to talke with a woman within the tearme of three yeeres, he shall indure such publibue Ferd. What say you Lordes? why, this was quite forgot. In pruning mee when shall you heare that I will prayse a hand, a foote, a face, an eye: a gate, a state, a brow, a brest, Ber. A toy my Leedge, a toy: your grace needs not feare it. Long. It did moue him to passion, & therfore lets heare it. Berow. Ah you whoreson loggerhead, you were borne to Ber. That you three fooles, lackt me foole, to make vp the Bero. True true, we are fower: will these turtles be gon? Clow. Walke aside the true folke, and let the traytors stay. King. What, did these rent lines shew some loue of thine? Ber. Did they quoth you? Who sees the heauenly Rosaline, Duma. Darke needes no Candles now, for darke is light. King. Then leaue this chat, and good Berowne now proue

This block of text is suitable for pasting into a word-frequency counter and thence into, say, a spreadsheet for analysis. Repeating the XQuery for full lines set by compositor T, for example, and counting the resulting words' frequencies enables rapid comparison of the kinds of spelling preferences that bibliographers are interested in. XQuery has rather more powerful features beyond the scope of this report, and they allow for example the extraction of lines in the order they were typeset (by specifying in the query the page order of setting) or just the lines on a particular inner or outer forme.

We may conclude with a survey of the limitations of the above approach. The greatest is that we cannot attach fresh attributes (such as statements about the printer's copy) to the individual XInclude lines in the documents representing the bibliographers' opinions: such attributes have to be either global for the compositor stint or encoded (globally or locally) into the base text. As displayed on the project website, the Internet Shakespeare Editions transcriptions of early editions have certain oddities, such as placing a turned-up line-ending below rather than above the line that it completes. The transcriptions correctly represent the early editions' habit of breaking a word across a line or an even page boundary, such as *elegance* beginning on E1v and ending on E2r in Q1 *Love's Labour's Lost*. Even if such a word were begun by one compositor and finished by another (which seems unlikely), for most analyses it would make little difference if we arbitrarily ruled that all

words belong with the page and line on which they began. XML documents may not contain lone ampersands as this character is reserved for special purposes; Q1 *Love's Labour's Lost* has dozens of them (standing for *and*) and they must be manually altered to XML's code for an ampersand.

Lastly, it may be objected that the present work fails to conform to the guidelines of the Text Encoding Initiative (TEI), which aims to provide an agreed standard for XML markup of literary works. Although TEI is acquiring techniques for representing physical documents – most notably the recently added guidelines for Genetic Editions – it has long privileged representation of the intellectual content of a work over representation of its material embodiment. The only existing TEI-conformant transcriptions of early editions of Shakespeare are those of the Text Creation Partnership (TCP) and those of the Modern Language Association's New Variorum Shakespeare. Both privilege the literary over the material form of the play: prose speeches, for example, are encoded as undivided paragraphs within <P> . . . <P> tags. The experiments described here suggest that one may quickly produce useful results by developing one's own XML conventions and applying them to readily available untagged electronic texts. This gives hope that alongside the large collaborative projects from which we have all been benefitting, including Internet Shakespeare Editions, the Text Creation Partnership, the New Variorum Shakespeare and the Shakespeare Quartos Archive, there remains a place in the digital future for lone scholars 'rolling their own' applications.

Footnotes

[1] Bowers, Fredson. *On Editing Shakespeare*. 2nd Ed. Charlottesvile: University of Virginia Press, 1966. 136.

[2] Sterne, Laurence. *The Life and Opinions of Tristram Shandy*. 2nd Ed., 9 vols. London: R. and J. Dodsley, 1761, 3. 169.

[3] Cummings, E.E. *Complete Poems 1904-1964*. Ed. George James Firmage.New York: Liveright, 1994.

[4] McGillivray, Murray. 'Ian Lancashire's Two Muses: A Belated Reply', in *Electronic Publishing: Politics and Pragmatics*, ed. Gabriel Egan, New Technologies in Medieval and Renaissance Studies (Toronto: Medieval and Renaissance Texts and Studies (MRTS) and ITER, 2010), 131.

[5] For some plays in the 1623 edition of Shakespeare (the Folio) the copy appears to have been manuscripts made by the professional scribe Ralph Crane, some of whose habits of spelling are known (T. H. Howard-Hill, *Spelling-Analysis and Ralph Crane: A Preparatory Study of His Life, Spelling, and Scribal Habits*, Unpublished PhD thesis, Victoria University of Wellington (New Zealand), 1960; T. H. Howard-Hill, *Ralph Crane and Some Shakespeare First Folio Comedies.* Charlottesville, VA: University Press of Virginia, 1972; Virginia J. Haas, 'Ralph Crane: A Status Report', *Analytical and Enumerative Bibliography*, 3, 1989. 3-10; The likely effect upon compositors' spelling habits of setting from such professional transcripts adds considerably to the difficulty of distinguishing their stints, Paul Werstine, 'Scribe or Compositor: Ralph Crane, Compositors D and F, and the First Four Plays in the Shakespeare First Folio', *Papers of the Bibliographical Society of America*, 95, 2001. 315-39.

[6] T. H. Howard-Hill, 'The Compositors of Shakespeare's Folio Comedies', *Studies in Bibliography*, 26, 1973. 61-106; T. H. Howard-Hill, *Compositors B and E in the Shakespeare First Folio and Some Recent Studies,* Columbia SC: Published privately by the author, 1976; MacDonald P. Jackson, 'Punctuation and the Compositors of Shakespeare's Sonnets, 1609', *The Library (=Transactions of the Bibliographical Society)*, 30, 1975. 1-24; MacDonald P. Jackson, 'Two Shakespeare Quartos: *Richard III* (1597) and *1 Henry IV* (1598)', *Studies in Bibliography*, 35, 1982. 173-90; MacDonald P. Jackson, 'Compositors' Stints and the Spacing of Punctuation in the First

Quarto (1609) of *Pericles*', *Papers of the Bibliographical Society of America*, 81, 1987. 17-23; MacDonald P. Jackson, 'Finding the Pattern: Peter Short's Shakespeare Quartos Revisited', *Bibliographical Society of Australia and New Zealand Bulletin*, 25, 2001. 67-86; Gary Taylor, 'The Shrinking Compositor A of the Shakespeare First Folio', *Studies in Bibliography*, 34, 1981. 96-117.

[7] Price, George R. 'The Printing of *Love's Labour's Lost* (1598)', *Papers of the Bibliographical Society of America*, 72, 1978: 405-434 (425).

[8] Paul Werstine, 'The Editorial Usefulness of Printing House and Compositor Studies: Reprinted from *Analytical and Enumerative Bibliography* 2 (1978): 153-165 with a New Afterword', in *Play-texts in Old Spelling: Papers from the Glendon Conference*, ed. G. B. Shand and Raymond C. Shady. New York: AMS, 1984. 35-64 (37-8).

[9] The present work was done on an office-standard laptop running Microsoft Windows version 7, ActiveState Incorporated's ActivePerl version 5.12, and SyncRO Soft Limited's XML editor Oxygen version 13.2, which incorporates the Saxon XQuery processor.