

“a creature native and indued/ Unto that element”? Digitizing *Hamlet*: How to Start; Where to Stop

Pip Willcox

Bodleian Library
University of Oxford, Oxford OX2 0EW
UK
pip.willcox@bodleian.ox.ac.uk

Abstract

Confronted with the task of publishing online 32 XML-encoded texts of *Hamlet* in quarto in the space of one year, the *Shakespeare Quartos Archive* (SQA) moved away from a traditional, editor-led approach to focus instead on end-users. This paper explores the process of the selection of and engagement with this audience, and assesses the effectiveness of this approach in terms of the digital editions, encoded to Text Encoding Initiative (TEI) p5 standard, and of the web interface the project created.

Background

The *Shakespeare Quartos Archive* (SQA) was a pilot project that ran from 2008 to 2009, producing digital editions of 32 copies of *Hamlet*.¹ Already digital editions of Shakespeare are rich and various,² but where the SQA differs from these other resources is in our engagement with users from the project's inception. Following the ideas and desires of these users has influenced the design and innovative functionality of the resource: it differs from other editions in its attention to the physicality of the books as well as to the texts they contain.

“Quarto” refers to a size of book,³ cheaper to produce than folios,⁴ exposing a publisher to a smaller financial risk in an uncertain market. 21 of Shakespeare's plays exist in quartos printed before 1642 (the significance of this date being the puritan-inspired closure of the theatres). The texts of these quartos are of particular interest because they represent the earliest surviving forms of many of the plays, predating the First Folio of 1623. For reasons whose causes are the subject of rich academic conjecture, the texts of the plays in quarto often differ widely from each other, as well as from those printed in the First Folio.

¹ To date, in addition to the web interface, the SQA has produced two book chapters, practical and theoretical descriptions, respectively, of our approaches to editing the texts (Siefing & Willcox, forthcoming; Kuhn, forthcoming).

² “Focusing just on sources for early printing”, Kuhn lists four digital resources (forthcoming). Add to the list the SQA interface, and the number of resources in this small niche alone equals the number of pre-1642 *Hamlet* editions. More appear on the SQA, <<http://quartos.org/info/links.html>>, and in Kuhn's paper, neither of which pretends to an exhaustive list, were such a thing possible.

³ The dimensions of a quarto can be up to about 24cm x 30cm (9 inches x 12 inches), depending on how much the pages are trimmed after printing. The name comes from the Latin *quartus* (fourth) as four leaves, each a quarter the size of the original, were produced from each single printed sheet, by folding it twice. (*Oxford English Dictionary*, s.v. “quarto, n. and adj.” Third edition, December 2007; online version September 2011. <<http://www.oed.com/view/Entry/156080>>.

⁴ A folio is a larger format, where single printed sheets are folded once, to give two leaves per sheet.

Hamlet, the play chosen for our pilot project, appears in five quarto editions before 1642, the first of which (printed in 1603) is just half the length of the play as it is best known today. It was chosen for the project partly due to this rich textual history, partly for its status as cultural icon, and partly because each of the six project-partner libraries⁵ holds at least one copy of the play.

Based on high quality digital images created by Octavo, a commercial publishing venture no longer in business,⁶ the SQA produced a digital interface that married the physical and textual elements of 32 copies of *Hamlet* in quarto held by six libraries.

This paper examines how the SQA project was modelled to answer the practical questions of text encoding and interface design arising from the decision to make the books' physicality paramount. Our answers included the consultation of stake-holders, the design of the XML schema, and the prioritization of certain interface features. This paper briefly describes the background to the JISC-NEH-funded collaborative project before addressing the methodology, implementation, results, and worth of the advice we sought from stake-holders.

Open Scholarship

As long ago as 1994, Robinson predicted that “there will be a place for ‘editions’ which are nothing more than compilations of materials, together with various tools for their exploration, just as there is now for series of facsimiles of manuscripts and for printed concordances” (1994). Putting to one side the question of agency elided by his “nothing more”, his sibylline statement begs: what tools do the putative users of such a site want? New to

⁵ The six participating libraries are the British Library, the Bodleian Library, Edinburgh University Library, the Folger Shakespeare Library, the Huntington Library, and the National Library of Scotland. Other partners in the project are the Maryland Institute for Technology in the Humanities (MITH), and The Shakespeare Institute of the University of Birmingham.

⁶ Some of the images of books taken by Octavo are available at <<http://rarebookroom.org/>>.

project modelling, it occurred to us that the most straightforward way to discover this was to ask them.

Advisory Forum

Defining (divining) who “they” were, inviting them, and negotiating a suitable date to meet was a venture in itself.⁷ A condition of the project funding was that the SQA interface be publicly available.

Given how widely accessible the internet is, our prospective audience was hugely diverse. We could imagine academic editors and researchers consulting the interface, teachers and lecturers introducing students to ideas of variant texts through the five quarto editions, typography and bibliography enthusiasts studying the digital facsimiles, directors and actors responding to the first printed version of *Hamlet*, codicologists viewing bindings, historians of book reception poring over the manuscript annotations... Perhaps primary school teachers would also show some of the less well preserved pages to children to encourage them to make tea- and turmeric-stained paper for homework, or students of book-binding would use publishing software to recreate the printing layout of the original sheets as they came out of the press, or collectors of library stamps or printed ephemera would paste copies of the marks of ownership and bookplates into scrapbooks.

Clearly we could not invite a representative of everyone who could feasibly use the web interface. Taking advice from partners and colleagues, we invited academic editors, researchers and teachers, education and performance specialists, and Text Encoding Initiative⁸ experts to our Advisory Forum. These Forum members were extraordinarily liberal in sharing their time and expertise, particularly when, in McGann’s words, they were colluding in the provision of the tools that would help “profoundly undermine” the “academic elite’s control over the circulation of knowledge” (2007). Perhaps, if there has ever been personal anxiety about this undermining, in the time between writing and our project, the usefulness of such freely available resources to members of the “academic elite” has overcome any perceived reservations about challenges to their authority such resources enable. Certainly other academics write in more positive terms about users becoming “collaborators in an ongoing creative dialogue” (Carson, 2008).

The day of the Advisory Forum comprised an introduction to the project and the sort of information we were hoping to elicit from the Forum members, break-out groups where people with common interests addressed particular issues, and a round-table discussion presenting their conclusions to the whole group and reaching a consensus of approach for the project. Forum members also completed a questionnaire which included a closed list of

⁷ Members of the SQA Advisory Forum, to whom we remain enormously grateful, are listed at <http://quartos.org/info/about.html>.

⁸ The Text Encoding Initiative (TEI) p5 was the XML standard we used to encode the digital editions of *Hamlet*.

potential features to be included in our editions and interface which they were asked to number in importance from one to ten.⁹

Amongst the lessons we learnt was that when faced with a simple either/or question (for example, “Obvious printer errors: record or correct?”), most people will answer “both”. Nonetheless, Forum members understood that in a time-limited project we could not encode every feature of the text in our digital editions, nor include every function a website might conceivably provide. It was instructive for us to see competing claims on project resources thrashed out. This process also crystallized the questions these interested parties wished a web interface could answer.

Researchers in typography were keen that every ligature (for example *ct*, *fl*) and every long *s* (*f*) be recorded, and that a precise measurement be noted of the distance between a mark of punctuation and the character preceding it. Editors wanting to navigate quickly through many quartos wanted spelling to be modernized, or a variant-spelling database to be included invisibly in the search facility. They also wanted act and scene breaks, absent in all the quarto editions, to be imposed on the texts.

The quantitative and qualitative data we gathered that day (and subsequently through the project’s message board and wiki) informed and elucidated our transcription and editing guidelines. It allowed us to take a view of where to start editing, informed by people who would use the interface, as well as giving us a clear point beyond which we should not go. With the guidance of our Forum members, we provided the encoding that allowed the web interface the functionality that they requested.

While it took significant preparation, collation and analysis, the Forum seemed an obvious and (relative to its usefulness) efficient means of defining a project model. Carson, a member of the Forum later described the process as “a very positive example” of a project “working to create networks of activities as well as resources” (2008). This became our guiding principle, that just as the project modelling had been a collaborative venture, so the web interface would allow, and indeed encourage, users to interact with the resource and the texts.

Editing

Our Forum members advised us about the sort of questions users might want to ask, and we kept this advice uppermost in our thoughts as we developed our transcription guidelines.

The guidelines derived from the Forum were constantly consulted as we balanced the limited time given to the pilot project with the level of detail of the books’

⁹ For more information about the Advisory Forum, including the questionnaire, see: <http://quartos.org/info/files/SQAAdvisoryForum.pdf>.

physicality and textual content that our XML-encoding should describe.¹⁰

As well as transcriptions of the texts of each of the 32 *Hamlets*, we included in the digital editions copy-specific details, implicit and explicit stories of the books' history, transmission, and reception. This information included manuscript annotations, binding descriptions, paste-ins, library stamps, book-plates and damage which rendered any of the text unreadable.

We discovered anew how simply many complex ideas can be transmitted, yet we found other aspects, straightforward in the originals, the most troublesome to describe in XML.

For example, manuscript footnotes to the printed text, labelled by line numbers (which otherwise do not appear in print or manuscript), and occasionally later cancelled are straightforward to encode.

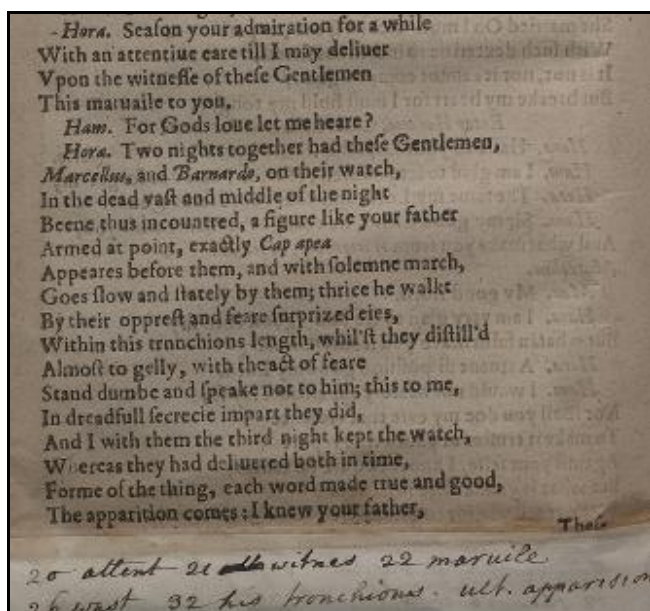


Figure 1: Complex annotations of the printed text (Bodleian Arch. G d.41)

The same applies to a printer's error, corrected in manuscript by over-writing the original.

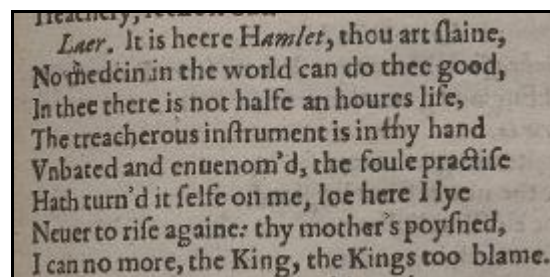


Figure 2: Manuscript emendation of the printed text (BL C.34.k.4)

But while the inclusion of hand-written annotations generally presented few difficulties, a speaker's name, apparently erroneously missed from the printed version (effectively putting a dialogue between Hamlet and the ghost of his father into just one voice) and added by hand in the margin part way (in terms of the *mise en page*) through the speech had implications which we simplified significantly in the encoding.

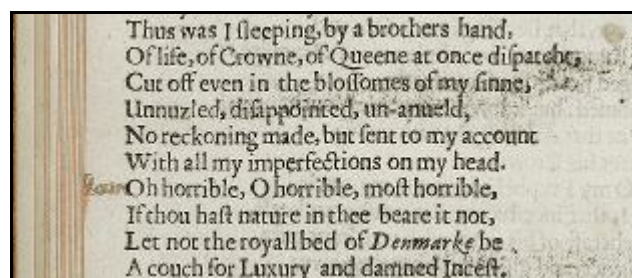


Figure 3: Manuscript division of a single printed speech (Folger STC 22279 Copy 2)

Our Forum responses were helpful too in stopping editing. We could have spent our lifetimes, as other people have, delving into the riches the text of *Hamlet* offers. While the creation of any edition is an act of interpretation, we concentrated on valid encoding that would enable users to pursue their own interests and research.

For example, we decided against measuring the distance between a character and the punctuation mark that followed it, while agreeing to record conventional act and scene breaks in the plays. In order to allow users to search for a word or phrase within a particular character's speeches, we recorded unique identifiers in the encoding around each speaker's name. This was especially important as some characters change names between the first and subsequent quartos, the most notable of which is that "Corambis" becomes "Polonius".

The Interface

The resulting web interface of the SQA encourages an interactive approach to the quartos. It brings together 32 copies of *Hamlet* in high quality images and searchable electronic text. Previously, even people fortunate enough to have access to one of the quarto-holding research libraries could never collect copies from across the United States and the United Kingdom in one room. It is therefore possible to view and compare textual variants

¹⁰ The transcription guidelines, with all the project documentation, can be found on the website, <http://quartos.org/info/documentation.html>. For further discussion of our editing methodology, see Siefring & Willcox (forthcoming).

directly, rather than reconstructing them through the critical apparatus of traditional scholarly editions.



Figure 4: The *Shakespeare Quartos Archive*, showing the title page of the first quarto edition of *Hamlet* (Huntington 69304)

In addition to this, the resource allows (but does not force) users to create an account where they can save versions of the SQA of their own. The tools included in the interface include the ability to crop, zoom and annotate the images, and save these in an “exhibit” for future reference, or to share with other users, or download for their own use.

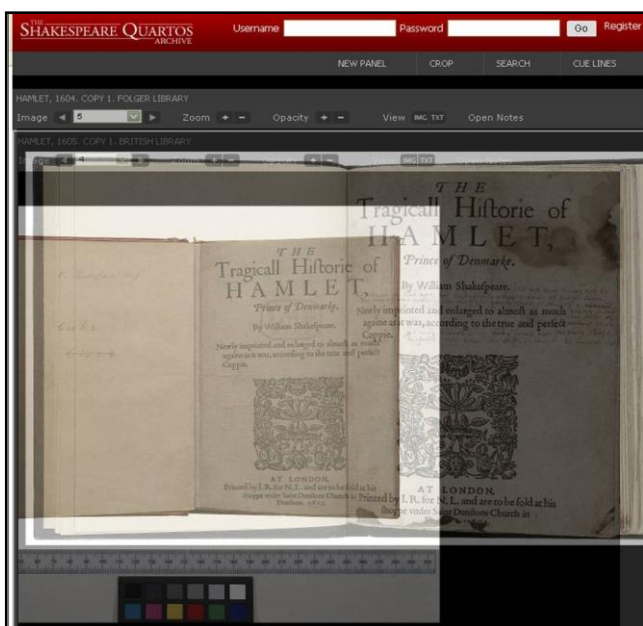


Figure 5: The *Shakespeare Quartos Archive*, showing two title pages of the second quarto edition of *Hamlet* (Folger STC 22276; BL C.34.K.2)

Multiple panels can be open at once, showing the digital editions and images. There is a tool for adjusting the opacity of each panel, giving users the ability to overlay semi-transparent images and function as an optical collator. Complementing this, is a textual collation tool

which makes use of the full-text transcriptions the project created.

There is a faceted search facility, so queries can be limited to one particular character’s speeches, or to one (or more) copies of the play. Responding to the wishes of users interested in history of the theatre, cue lines and speeches for any single character can be easily produced from the transcriptions.

Not only is the interface freely available, but also the software and tools developed for the project, the images, and the XML editions can be downloaded under a Creative Commons license.

The Future

The SQA succeeded in creating an interface to support research and discovery, continuing the “iterative dialogue” with our users (Carson, 2008). Reviews have been positive: for example, it has been described as “useful, thoughtful, and efficient” (Trettien, 2010). What is missing from a project such as this is follow-up. At a local level this means that, despite having created a two-year project plan to extend the treatment to the 20 other plays printed in quarto before the 1642 theatre closure, the pilot *Hamlet* seems set to remain the only play in the archive.¹¹

On a more general level, an accidental effect of the proliferation of soft-funded digital projects is that there is rarely the time (even where the personnel are still present) to write retrospective reports: how many people use a resource; who uses it; how do they use it; what changes or improvements do they envision; how sustainable is the resource proving? Coupled with the perceived need for positive project write-ups, one wonders how many times mistakes learnt in one project are needlessly duplicated. As a community, we need to learn that failure can be the parent of success, and might usefully be celebrated as such (Harford, 2011).

Forums such as ours, and such as this, where we question, respond to and discuss our work, its practical, and its theoretical implications are vital, breaking down divisions between the people who write about and the people who practise digital humanities. If we are to support digital scholarship, offering a readership access to special collections and specialist tools “with a convenience never before available” (Robinson, 1994), enabling “someone else” to “do the coolest thing” with our data (Pollock & Walsh, 2007), we need to ask more questions, and listen to the answers.

Acknowledgements

The project which is the subject of this paper could not have happened without the generous funding the JISC-NEH Transatlantic Digitisation Collaboration Grant, to which I am very grateful. Details of the first round of this

¹¹ A list of the quarto plays appears on the British Library’s partner website, *Shakespeare in Quarto* <<http://www.bl.uk/treasures/shakespeare/quartos.html>>.

funding programme in the UK can be found at the website of the Joint Information System Committee (JISC), <http://www.jisc.ac.uk/fundingopportunities/funding_calls/2008/12/grant1308.aspx>. JISC's counterpart in the US, The National Endowment for the Humanities (NEH), details its administration of the programme here: <<http://www.neh.gov/grants/guidelines/JISC.html>>.

The project flourished thanks to the resources and expertise of colleagues at all the participating institutions: the British Library, the Bodleian Library, Edinburgh University Library, the Folger Shakespeare Library, the Huntington Library, Maryland Institute for Technology in the Humanities, the National Library of Scotland, and The Shakespeare Institute of the University of Birmingham.

In particular I am indebted to Jim Kuhn, of the Folger Shakespeare Library, and Doug Reside, formerly of MITH and now at the New York Public Library, and to my immediate colleagues at the Bodleian Library, Simon Charles, Emma Huber, Colm MacCrossan, Richard Ovenden, Michael Popham and Judith Siefring, for their generosity and enthusiasm.

References

- Harford, T. (2011). *Adapt: Why Success Always Starts with Failure*. London: Little, Brown.
- Carson, Christie (2008). Shakespeare Online: an increasingly interesting conversation. In O. Terris, E. Oesterlen, & L. McKernan, (Eds.), *Shakespeare on Film, Television and Radio* (pp. 94--112). London: British University Film and Video Council.
- Kuhn, J. (forthcoming). "A hawk from a handsaw:" Collating Possibilities with The *Shakespeare Quartos Archive*. In Tassie Gniady, Kris McAbee, & Jessica Murphy (Eds.), *New technologies in Renaissance studies*. Tempe, AZ: Arizona Center for Medieval and Renaissance Studies.
- McGann, Jerome (2007). Database, Interface, and Archival Fever. *Publications of the Modern Language Association*, 122, 5 (10), 1588--1592.
- Oxford English Dictionary. <<http://www.oed.com>>.
- Pollock, R. & Walsh, J. (2007). Atomisation and Open Data. XTech Day 3. <http://blogs.talis.com/nodalities/2007/05/xtech_day_3_rufus_pollock_and_php>.
- Rare Book Room. <<http://rarebookroom.org/>>.
- Robinson, P. M. W. (1994). Collation, Textual Criticism, Publication, and the Computer. *Text*, 7, 77--94. <<http://www.jstor.org/stable/pdfplus/30227694.pdf>>.
- Shakespeare Quartos Archive. <<http://quartos.org/>>.
- Shakespeare, William (1623). *Mr William Shakespeares Comedies, Histories, & Tragedies*. Published according to the True Originall Copies. London: Isaac Jaggard and Edward Blount. <http://eebo.chadwyck.com/search/full_rec?SOURCE=config.cfg&ACTION=ByID&ID=99846615>.
- Siefring, J. & Willcox, P. (forthcoming). More than was Dreamt of in our Philosophy: Encoding *Hamlet* for the *Shakespeare Quartos Archive*. In Brent Nelson & Melissa Terras (Eds.), *Digitizing Material Culture*, from

Antiquity to 1700. Toronto & Tempe, AZ: Iter/Arizona Centre for Medieval and Renaissance Studies.

Trettien, Whitney Anne (2010). Disciplining digital humanities, 2010: Shakespeare's Staging, XMAS, Shakespeare Performance in Asia, Shakespeare Quartos Archive, BardBox. *Shakespeare Quarterly* 61 (3): 391--400. <http://muse.jhu.edu/journals/shakespeare_quarterly/v061/61.3.trettien.html>.