

Approaches to digital editing*

The editing and circulation of historical texts has long been a part of cultural activity in many societies and parts of the world, as a means both to preserve and interpret the past. However, while editing schemes and standards vary between authors and publishers, the nature of this scholarly activity has changed relatively little over the centuries, mainly because of the limitations of print-based modes of production. In the digital era some progress has been made by the humanities: within the editorial world, for instance, there has been consideration of the use of virtual workspaces, although their realisation is still some way off. Similarly, there are tools available which, to some degree, automate the processes of text collation and comparison. Yet the actual processes and methods of text editing have remained essentially the same.

Digital editing promises to change all of this, and in so doing help scholars to rethink their relationship with the texts themselves. Editing may be collaborative, either between individual scholars or with the involvement of the wider community. It can take place virtually and remotely. The text may change over time, with novel insights and interpretations being used to create new versions to compare with the old. Publication methods may become more varied and flexible. Texts may be edited within online research environments, creating dialogues or links between them and other source materials. There is ultimately the potential to transform what we conceive to be an 'edition', but in order to achieve this, new academic tools have to be developed, and scholars need to be both involved in their creation and supported in their use.

Quite apart from the exciting possibilities afforded by digital editing tools and technologies, there is a rather more prosaic imperative at work, that is, the increasing unsustainability of current print publishing models. This is particularly the case with regard to large-scale editions and reference

* This paper was delivered at the 'Envisaging REED in the digital age' workshop, organised by the Records of Early English Drama project at the University of Toronto 4–5 April 2011. Thanks are due to the organisers, Sally-Beth MacLean and Jason Boyd.

works, where researchers are faced both with demonstrably successful online-only resources and an increasing reluctance on the part of academic presses to take on the publication of scholarly editions.

So what is the way forward? As funding for the digital humanities diminishes, and this is likely to be the case for the foreseeable future, there is a growing emphasis on the development of both central platforms, which may be utilised by a range of projects, and open source tools, which may be created in the context of a specific project but can easily be repurposed and reused. Perhaps related to this financial constriction, but certainly to lessons learnt from previous failures, there is an encouraging move towards greater consultation with researchers at an early stage in the creation of such resources. Too often in the past the technology has driven the change – what we can do rather than what we should – but this does not seem to be the case with regard to digital editing.

I will be discussing three projects today which are of potential relevance for Records of Early English Drama (REED) online, and in all cases academic considerations have been central to their development. The first, ReScript, is a project of the Institute of Historical Research which aims to build a prototype editing platform which may be used in the creation of different types of edition, employing very different workflows.¹ Central to the project has been a survey of requirements. An online survey, conducted at the beginning of the year, received just over 1,000 responses and demonstrated a significant degree of interest in the transcribing and editing of texts, both from within and outside formal higher education. In addition, a series of interviews is underway, the aim of which is to establish which elements of the editorial process are common across all time periods and types of text, and consequently may be provided at a central level with no or minimum customisation between projects.

¹ <http://www.history.ac.uk/projects/ReScript> [accessed 7 April 2011]; and see the prototype at <http://rescript.org/Default.aspx> [accessed 7 April 2011].

The prototype will be tested using three texts, representative of different approaches to historical editing, in order to ensure that it is sufficiently flexible to cope with the range of editorial activity undertaken by historians. The *History of the Rebellion*, written by the Earl of Clarendon in the 17th century, consists of around 1.5 million words and is a composite of two main texts and many smaller ones. It will be produced by a team of editors based in various universities and research organisations across the UK. The St Botolph's Aldgate Parish Memorandum Books, produced between 1586 and 1625 are less complex, but run to 2,584 folios (just over 1.3 million words). The edition is being produced by a small research team located within the Centre for Metropolitan History. Finally, it is planned to produce a new edition of Joseph Foster's *Alumni Oxonienses*, published between 1888 and 1892. The entries in this biographical dictionary are relatively sparse, and the potential to enhance them, including by linking to other online resources such as the *Oxford Dictionary of National Biography*, is enormous. Rich in personal names, the *Alumni* is an ideal candidate for crowdsourcing, and the platform will be developed to support this much more unstructured approach to editing.

While this is, of course, no substitute for a scholarly edition, it is an option available once editing takes place in a digital context. Contributions of this kind can be subject to strict quality assurance and form part of a final edition, or they can be unmediated, and presented entirely separately, with appropriate health warnings. There have been a number of successful examples of crowdsourcing in recent months. The excellent Transcribe Bentham project at University College London has perhaps been the most widely publicised. To date just under 400 individuals (as opposed to 1,200 registered users) have been responsible for producing over 1,000 transcripts. The software used for the project is open-source and consequently may be re-used or adapted by other editing initiatives. The Transcribe Bentham team readily admit that a full-time postdoctoral researcher would have made a greater contribution over the life-time of the project, but crucially the model they have chosen has the potential to allow for work to continue, albeit on a reduced scale, once funding has ceased.

Other tools are already available, for example the recently launched Scripto, which includes options for version control (and history) and different levels of editorial control.²

We are still at a relatively early stage in the development of the ReScript platform, but a rudimentary interface has been set up to allow our editors to begin to experiment with what's possible, and to feed their requirements back in to the design process. In this case the printed text of the *Alumni* has been double re-keyed and marked up for people, places, occupation, and so on. The very structured nature of the text allowed most of this process to be automated, with only spot checking for accuracy. For example, each entry begins with the form 'Surname, Forename'; any word which precedes the abbreviation 'Coll.' will refer to the Oxford college attended by the individual in question. It's far from perfect – you'll notice problems with some of the spacing here – but has been achieved in only three months. You can begin to see some of the possibilities offered by creating and presenting texts in this way from the filtering options on the left. For example, clicking 'vicar' in the list of occupations recorded brings up 231 records, which can be further filtered by clicking on the college or some other element which will be marked up in due course.

Turning to a single item within the edition, the entry for Bartholomew Abbott, you can see the type of mark-up that has been applied: forename and surname (within the class of person), college, status, age at matriculation, and so on. This is clearly not the easiest view of the material to work with, so a basic XML editor has been included. A lot of the options at the top here will be removed, and the 'quick formatting' bar on the right either replaced with or supplemented by the semantic mark-up appropriate for this particular text, but the general idea remains the same. An image viewer, which allows editors to consult a version of the original document alongside the edited text, will be the next element to be added.

² <http://scripto.org/> [accessed 7 April 2011].

ReScript is deliberately not intended to be a publishing platform; rather it will offer a range of export options which will support publication elsewhere, whether on an existing project website, as an ebook, within an established digital library, or for print. Consequently, editions produced using the ReScript service will not be dependent on a single platform for their communication and dissemination and can be repurposed as necessary.

The second project that I'm going to talk about is Early English Laws, a collaboration between the IHR and the Department of Digital Humanities at King's College London.³ It received an initial three-years' funding from the UK's Arts and Humanities Research Council, but is conceived as a much longer, 10-year project. The aim of Early English Laws is to edit or re-edit and translate more than 150 early English legal texts, providing each with an introduction and full commentary. It has been a complex project, and most of the first two years was taken up with sourcing the manuscript images, held in libraries and archives throughout Britain and Europe, and with an analysis of the texts. Crucial here has been developing an interface which will be able to cope with the intricate connections between the texts, many of which exist in multiple versions, in different languages, and with a great deal of borrowing between law codes over time.

Having recently completed the analysis, we are now beginning to construct the editorial interface, which, like the Transcribe Bentham project, will be made available as open source code next year.⁴ It is inevitably a bespoke product, designed to cope with the vagaries of this particular collection of material, but it could be customised relatively easily for use by other projects which are handling a diverse corpus of material. You can see from the main administration screen that there are a large number of elements which need to be accommodated in relation to each text, from details of the archives in which related manuscripts are held to information about the different witnesses that are

³ <http://www.earlyenglishlaws.ac.uk/> [accessed 7 April 2011].

⁴ The interface has been developed by the team at the Department of Digital Humanities, Geoffroy Noel, Eleonora Litta Modignani Picozzi, Elena Pierazzo and Paul Spence.

still in existence. A level below this, each text has its own home page as it were, where basic metadata may be added and from which the editor or editors can access the translation, transcription and so on. Turning to the editing process itself, you can see that we have progressed further than with ReScript. We have, for example, begun to add navigation within the text (on the left here), taxonomies and so on. While there is significant variation between texts, they are all structured in a similar way, with a prologue and division into books, chapters and sub-chapters. A sub-chapter is the smallest textual element that we have needed to mark up structurally, but this could easily be expanded should additional layers be required either by this or by other projects.

At the top of the XML editor, there are buttons to identify an area of text as a particular type of division, in this case a book, and it is also very easy to mark-up text which has been interlined, deleted, and so on. Here, a word – ‘iusticiam’ – is highlighted in the text and then a single click marks it up as supplied by the editor. We have devised authority lists of people, places and particular legal terms or concepts, and words and phrases can be marked up in the same way using the drop down list at the top. This will allow us to filter out all instances of, for example, King Alfred across all of the texts included in the database. The mark up is clear without having to engage with the underlying coding, although this can be revealed by clicking on ‘html’ at the top of the editor. You should just be able to see that we’re using the Text Encoding Initiative (TEI), which is the best means of ensuring interoperability further down the line.

The left-hand menu essentially works as a table of contents for the manuscript being edited. You can choose to view the whole document, with the individual divisions clearly visible, or work on a single sub-division. The editor can clearly see where in the text he or she is working, as it will be highlighted in bold in the navigation structure. At the moment this is very clean and clear, but we will need to find a way of adding in a panel for the critical apparatus, as well as an image view which will make available a facsimile of the original manuscript. This will most likely be in a pop up window,

but we'll have to do some more user testing before we decide on the optimum arrangement of the different elements.

As with the *Alumni* text that you saw earlier in ReScript, it is essential that the text being edited is divided up in some way, and ideally into relatively small sections. Practically, this will speed up the work of editing, as an entire document is not being saved every time an editor makes a small change to one element. But there are real issues of functionality as well. In the case of Early English Laws, we want to be able to publish alongside each other, in various combinations, the manuscript images, the edited text, the critical apparatus and the translation of the document, at a minimum (there may for example be a requirement to allow the display of medieval Latin and French translations of the original Old English as well). In order to do this, it is essential that a section of text links to a folio image, and then on to a footnote and a particular point in the commentary. Breaking the text up into manageable elements is the only way of handling this, or at least the only way in which we have been able to do it.

While we have a project website for Early English Laws, which has short descriptions of all the texts and a variety of other information, we have not yet developed the publishing front end. However, the approach will be similar to that adopted by the team behind a new online edition of the 902 surviving letters of Vincent Van Gogh.⁵ You can see here that the default view is to present the original text on the right alongside the English translation. A smaller central column contains information about the letter including location, date, subject and source status. The same display options are available in each of the main columns, allowing users to toggle additionally between the text with original line endings preserved, a facsimile of the letter, notes (which you can also choose to see individually in the central column) and any sketches included by the artist. This is a very impressive site, and one of the best presentations I've seen of such complex material. The

⁵ <http://vangoghletters.org/vg/> [accessed 7 April 2011].

production methods they've used have also incidentally supported the publication of an accompanying lavish print edition.

The final initiative that I'd like to mention today is the German TextGrid project.⁶ It is described as a Virtual Research Environment for philologists, linguists, musicologists and art historians, but the applications are potentially much wider. It is explicitly concerned with handling text rather than, say, research data, and has been developed in such a way that it should be relatively easy to plug in tools, for example dictionaries and bibliographic resources, which would allow it be used by researchers in other humanities disciplines. TextGrid is currently only available in a beta version, but I was involved in some user testing a year or so ago and it looks very promising indeed.⁷ It already has many of the functions that you would require from a collaborative editing platform, including an XML editor; a text-image link editor, which allows connections to be made between sections of text and specific points on an associated image; a metadata editor which supports searching across different projects within TextGrid; a customisable workflow tool, and so on. It's available for free download and I would certainly recommend exploring it.

In conclusion, there is already a great deal of work going on in relation to digital editing, and to virtual research environments for the humanities more widely. Some of these initiatives are already well advanced; others are still at a stage where their development might be influenced by the requirements of an established project like REED, or at the very least benefit from consultation with such an experienced team of editors who are already thinking about many of these issues. Almost all of this work is being conducted in an open-source context, with the intention that it will be adopted by other researchers and then customised and improved for the benefit of the community as a

⁶ <http://www.textgrid.de/en/ueber-textgrid.html> [accessed 7 April 2011].

⁷ This was as part of the TextVRE project at the Centre for e-Research, King's College London. See http://stuardunn.files.wordpress.com/2010/06/eel_usecase_1.pdf for the Early English Laws use case produced by Stuart Dunn.

whole. Finally, it is taking place in an altered funding environment, with both public bodies and private philanthropic organisations looking favourably upon collaboration and the re-use of existing best practice. All in all, it's a very exciting time to be an editor!