Ownership, Institutions, and Methodology

Mussell, James

Link to publication on Research at Birmingham portal

Ownership, Institutions, and Methodology

James Mussell

Ownership, Institutions, and Methodology

James Mussell

My own area of research, the study of nineteenth-century serials, has been and will continue to be radically transformed by the impact of digital technologies. Whereas before periodicals were the main objects of study for a relatively small group of scholars, and were treated as a source of background information for most others, the digitization of the nineteenth-century print archive, through the simple fact of making this material more accessible, has the potential to return the periodical to its central place in nineteenth-century studies. Whereas previously the study of periodicals meant recourse to indices and hours over bound volumes in libraries, it is now possible to search and read a vast range of titles from your desktop. However, the proximity of the nineteenth-century periodical archive to nineteenth-century research is not enough: without developing corresponding methodological approaches in how to think about and use these resources, we remain trapped in methodologies shaped by our encounters with certain forms of printed objects in certain dusty rooms.

The amenable condition of the periodical press for digitization – there is lots of it; it is largely out of copyright; substantial swathes of it are available on microfilm; there are demonstrable profits to be made – has lead to a situation where there are, at present, three large digital projects, all of which are producing large resources of nineteenth-century periodicals. Two of these resources are being produced by private companies, Proquest and Gale, and the third is by the British Library (but distributed by Gale). Of these, JISC (the Joint Information Systems Committee) will provide free access for to the British Library project for HE institutions, while access to the other two is by subscription from Proquest and Gale respectively.[1] This immediately raises the question of ownership: as the material that is digitised is in the public domain and looked after by public institutions, should the public have to pay again for access to it? Scanning material and constructing appropriate user interfaces and data structures is labour intensive and expensive. The costs of digitization must be met somehow, and subscriptions allow the large sums involved to be worked in to more modest institutional budgets. Indeed, the costs involved are so high that the academic sector, which usually makes its digital products available free online in exchange for public money, is largely priced out of this market.[2] Instead, the pattern of granting one-off awards over fixed periods of time to projects with definite deliverables has produced, as Julia Thomas and Jerome McGann note, a host of

well-designed projects that all stand independently of each other.[3] This, in turn, has encouraged research into creating networks that can bring together these projects, providing portals that can search across diverse contents, and preserving the data against the vagaries of institutional politics.[4]

One place where issues of ownership are focused is in the corporate battle between Microsoft and Google over their respective search engines. Google's controversial Google Book Search scanned the contents of books sourced from libraries and publishers in order to create an index from the OCR-generated text they produced. This index then becomes part of Google's project to 'organise the world's information and make it universally accessible and useful', providing access to the full text if the book is out of copyright, and to snippets of the text from search results if still in copyright.[5] However, what this means in practice is 'universally accessible' through Google's search engine, and useful within the scope of their treatment of the work when digitised. Microsoft's response has been to set up a rival project in collaboration with the British Library, the University of Toronto, University of California, New York Public Library, American Veterinary College Library, Yale, and Cornell. The institutions are each responsible for providing a section of their holdings according to Microsoft's themes: the British Library, for instance, is responsible for 'English Literature' and it has used its own classification scheme to identify the material that this entails. Like Google, Microsoft are locking partner institutions into a license agreement that prevents them making the digital content that is produced accessible in any way other than through an MSN search. Both of these projects are immensely useful, particularly for material that is out of copyright, as they will make the contents of texts searchable while providing access to millions of facsimile page images. However, they are underpinned by a tawdry deal in which the custodians of the culture of the Western world hold the material with which they are entrusted to ransom (for the period of a license) so that Microsoft and Google can sell advertising.

It is common for large-scale digital projects to consult with the academic sector when considering the scope of their projects – after all, the sector (including its students) is the primary market for such resources – but very little research is undertaken in the design of the projects themselves. Large scale resources, driven by the logic of the market, are inherently conservative as they will only be innovative within certain narrow constraints in order to beat their rivals, and will certainly not engage in riskier activities such as designing resources

that allow users to think about material in novel ways. The logic behind the design of user interfaces is that of the culture of web design more broadly: point and click, with navigation made as simple as possible; or, as the title of Steve's Krug's influential book has it, *Don't Make Me Think*.[6] Such simple design is deceptively difficult, and designers are skilled practitioners who take into account detailed user studies. The sums involved in producing digital resources are so large that the corporations and institutions that fund them attempt to guarantee that they will be used by enough people so that, in turn, libraries will be under pressure to purchase subscriptions. The result is well-designed projects, that do not need a great deal of expertise to use but might have more advanced features tucked away within them. Although often extremely powerful, these projects do not apply rigorous scholarly care to the material that they contain, and they do not risk alienating the majority of their prospective users by presenting content in a way that demands we ask new questions about it.

The market, then, produces resources that are within established patterns of use: although the rhetoric is of providing access to rare materials, and permitting them to be searched (often for the first time), the way in which this is accomplished is through minimal attention to the materials themselves. In terms of serials, these resources facilitate well-established prejudices about the genre. Despite the well-recognised importance of the periodical press in the nineteenth century, the archive is still predominantly used by scholars to provide background to more familiar works from the period. Digital resources, because of the way they have been designed, facilitate the mining of their contents for information, and so will consolidate the idea of the press as an undifferentiated mass of content from which to abstract choice portions. Despite the differences in user interfaces and the level of indexing and metadata involved, the model remains that of Google's search engine: a simple search field that produces a vast number of results through which users must browse.

This will have an important impact on the field. By making hundreds of periodical titles so accessible, these projects will transform nineteenth-century studies by recontextualising familiar writers, texts and events, and populating the field with references to otherwise obscure works. In periodical studies this will alter the way we privilege certain titles as exemplars – for instance *Punch* as voice of middle-class opinion; *The Times* as shorthand for the daily press – by providing equally accessible counter-examples. It will also turn up all sorts of interesting anecdotes and scraps of information, some of which might change the way we think about aspects of the period. However, using

these resources will encourage a methodology that depends upon the results of searches that users do not understand. Rather than read tracts of the press, users will instead encounter spots of it – usually articles – that are not only divorced from the other articles on a page, but also from the section, issue, and volume that they are within. Although these methodologies reproduce the serendipitous nature of more traditional forms of research, encounters with articles abstracted from their immediate print context makes it harder to establish connections between them and, without turning the pages, it is much more difficult to find the things that one did not know were there. Long lists of search results give the impression of exhaustiveness but, as Patrick Leary has argued, browsing search results is not the same as reading whole pages.[7]

In other words, what the digitization of the nineteenth-century periodical archive will not advance is people beginning to take the periodical press seriously as a medium in its own right. As students and scholars will increasingly use material from periodicals to support their arguments, it becomes, paradoxically, increasingly important that they appreciate what it is that is represented on their screens. The way to do this is to use the accessibility of these resources to bring them into the classroom. The teaching of the periodical press has always been rendered difficult by the fragmentary, dispersed and damaged condition of surviving archives. Even in institutions where there are substantial runs of serials, often readers cannot remove them from the library and there is rarely more than one copy of a particular issue or volume. Using digital resources removes the need for damaging practices such as photocopying as images can be printed out or, if facilities allow, projected onto a wall or read on individual monitors in class. For the first time it becomes possible to set reading directly from the periodical press and, if access to resources is permitted remotely, students can be encouraged to explore the abundance of material within the archive on their own.

As students will come across material from periodicals and news-papers during their own reading anyway, it is important that serials are covered in all courses on nineteenth-century culture. The diversity of material within the press and the different ways in which it is presented require a broader approach to the period. For instance, students in literary studies will need to think about anonymous writing, writing for commercial ends, and the situatedness of all contributions to the periodical press. Equally, the use of page facsimiles will draw attention to the way pages look as well as what they say, broadening the notion

of reading to include visual components of texts such as typography and, as Julia Thomas reminds us, images. Perhaps most challenging – both to students and their teachers – is being confronted by material that they do not know. It has always been the case in periodical studies that the amount there is to read surpasses the reach of any one reader. The possibility that a telling counter-example to an argument will appear on the next page is both exciting (it keeps us turning the pages) and disconcerting as it reminds us that much of what we know remains provisional. In an environment where material is more readily searchable while also being searched by more people, it will become difficult to make definite claims about aspects of the press based upon narrow readings of case studies, and students will always be able to bring material to class that their teachers have not encountered before.

   Although teaching nineteenth-century serials is important, it is also important to get students to think critically about how such material gets from hard copy to their screens. Students and staff already rely on a range of digital technologies within education, but these resources are largely 'black-boxed', with little understanding of either how they work or how they are produced.[8] This can produce a peculiar situation where, for instance, close readings are done on e-texts that look nothing like the works from which they are produced. The suggestion that there are some aspects of form that belong to the text (rhyme, punctuation) and others that do not (typography, layout), reinforces the notion of text as transcendent and reproducible at the expense of materiality and context. As the former properties are easily reproduced digitally, this notion also permits the assumption that digital resources can provide surrogates for published works without considering what is lost in their production. It is important that we recognise the editorial work that goes into producing digital resources and that we think seriously about the various transformations that material must undergo in order to be delivered on screen. We are at a moment when using digital resources still seems relatively novel and so we should use the sense of estrangement we often feel when we encounter a text in digital form that we know well in hard copy to ask questions about why it seems different. Rather than simply deploring what is missing (the smell of mouldering paper, the texture of a page), we should also recognise what we gain (searchability, access). It is only by understanding the relationship that digital editions have with the objects from which they are produced that it is possible to critically engage with resources while recognising the work that goes into creating them.

Institutionalising digital humanities is both a means of entering debates about the ownership of electronic media while also ensuring that we have the skills to critically evaluate the resources that are produced. In the slow-moving world of academia there remains a reluctance to take digital resources seriously as forms of scholarly production despite the fact that they have been appearing for almost twenty years and are widely used throughout the sector. We still consider the journal and the monograph as the marketable (and, in the light of the prospective metrics assessment regime in the UK, increasingly quantifiable) markers of academic expertise; we still expect doctoral students to produce theses that nobody reads. Funding bodies have recognised their role in resourcing the production of digital resources, and the increasing number of postdoctoral opportunities that such projects provide have not only provided valuable support for those scholars who have recently finished PhDs, but have also equipped them with a host of new skills. However, with the AHRC's withdrawal of funding from the Arts and Humanities Data Service (AHDS), there is a danger that such skills tend to be lost as projects come to an end and those working on them move to new posts elsewhere. Although some projects receive more use than others, the potential audience and impact for these resources far exceeds that of the academic monograph. Yet reviewers are reluctant to review digital resources alongside more traditional research outputs, and appointment committees will prefer a book or a journal article over experience of working on a digital project. Preparing nineteenth-century material for digitization requires that we reconsider how we conceptualise it. If we continue to see such work as separate from the business of being an academic in the humanities, we close our disciplines both to the experience of those who have carried out research in the process of producing digital resources while also failing to take the resources seriously as objects of study in their own right.

This means that serious attention must be paid in both teaching and research to the unglamorous aspects of digital production such as how data is encoded and organised, markup schema, the accuracy of OCR transcripts, and the scanning technologies that make images look as they do. The history of the book, of course, has been directing scholarly attention at aspects of textual production such as sales figures, paper quality and printing which are equally as unglamorous as the use of xml. The digital humanities belong to a much broader history of knowledge production that also encompasses the history of the book. Producing digital resources from nineteenth-century material requires us to ask new questions about it in order to define what it is we want

to reproduce and what it is that we can bear to lose. As the resources we create become the resources that scholars and students use, it is essential that such choices are available for critical scrutiny. As those working with nineteenth-century periodicals are all too aware, it is only by understanding how works are produced – whether on canvas, stage, paper or in electronic form – that we can understand why they appear as they do while remembering that they might still appear in a different form elsewhere.

<div align="right">

(University of Birmingham)

</div>

Endnotes

1. JISC have recently undertaken a consultation process to determine whether they should negotiate a license for the Gale project to allow HE institutions to subscribe to it at a much-reduced rate.
2. Notable exceptions in the UK include the Darwin Correspondence Project, Darwin Online and the Newton Project. It is telling that all of these long-running projects are focused on the works of individual famous Englishmen. See http://www.lib.cam.ac.uk/Departments/Darwin/; http://darwin-online.org.uk/; http://www.newtonproject.sussex.ac.uk. For a review of Darwin Online, see this special issue.
3. See also Jerome McGann, 'The Way We Live Now: What is to be Done?' http://www.nines.org/about/bibliog/mcgann-chicago.pdf.
4. See for instance JISC's Portals Program, the useful attempts by the Arts and Humanities Data Service (AHDS) to aggregate experience and skills (which will cease in April 2008 as a result of the depressing decision by the Arts and Humanities Research Council to remove its funding); and the Networked Interface for Nineteenth-century Electronic Scholarship (NINES). See http://www.jisc.ac.uk/whatwedo/programmes/programme_portals.aspx; http://ahds.ac.uk/; http://www.nines.org.
5. See Google's own account of the production of Google Books Search at http://www.google.com/googlebooks/newsviews/history.html.
6. Steve Krug, *Don't Make me Think: A Common Sense Approach to Web Usability*, 2000 (Indianapolis: New Riders Press 2006). The book exemplifies research into new media and the attention economy. See for instance Thomas H. Davenport and John C. Beck, *The Attention Economy: Understanding the New Currency of Business* (Cambridge, MA: Harvard Business School Press 2001).
7. Patrick Leary, 'Googling the Victorians', *Journal of Victorian Culture*, 10 : 1 (spring 2005), 71–86.
8. The term 'black box' is used by Bruno Latour to describe how a scientific concept or technology becomes accepted as a discrete object in society rather than an assemblage of the various constituent networks from which it is constructed. See Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society* (Cambridge, MA: Harvard University Press, 1987).