# 9

# Digital Humanities and the Future of the Book

*Marilyn Deegan and Matthew Hayler*

## A brief history of digital humanities

Having been a somewhat niche activity for decades, digital humanities (DH), formerly called humanities computing, leapt into prominence in 2009 when it was pronounced the 'next big thing' at the US Modern Languages Association conference. But what is 'digital humanities'? In a world where there is surely no one in the humanities who doesn't use digital tools and resources, is digital humanities something special? There are specialist journals, collections of essays, and monographs devoted to it. There are also departments and centres of digital humanities in many institutions, and job lines in English (and sometimes other disciplines) and digital humanities. The US National Endowment for the Humanities has an Office of Digital Humanities, the UK's Arts and Humanities Research Council has a Digital Transformations theme, other funders eagerly accept proposals in the digital humanities. There seems to be a prevailing view that, as Parker points out to us, 'project plus digital equals funding' (Parker, 2012, p. 3). What's going on?

In this chapter we investigate a number of the important events in the history of what the Digital Humanities might be; the technologies that have underpinned its advances and presented its challenges; and some of its particular trajectory in English Literature departments. The chapter also ends with a short case study to interrogate the boundaries of DH, asking what should be left out of the field. Whilst we cannot be comprehensive here, we hope to outline the most significant events that have led to DH's current manifestation. Our focus on English Literature principally stems from the way in which the field has become embedded in such departments, and their tendency



A. Hewings et al. (eds.), *Futures for English Studies*

towards capacious understandings of reading (of which more below), and our own institutional backgrounds. From its inception, however, DH has been a multidisciplinary affair, led, at various times, by Computer Science, Classics, Religious Studies, and English Language and Linguistics. Each of these fields continues to shape our understanding of the potential for contemporary research in the Humanities, but this chapter isn't intended to be the final word on what DH is and has been, instead charting one path for those new to the field and interested in where it might go.

In 2010 Matthew Kirschenbaum, himself a scholar of English working in the US, wrote a blog post called 'What is digital humanities and what is it doing in English departments?' (Kirschenbaum, 2010). This was widely circulated and has appeared in at least two collections of articles on digital humanities. His definition of digital humanities is:

> The digital humanities, also known as humanities computing, is a field of study, research, teaching, and invention concerned with the intersection of computing and the disciplines of the humanities. It is methodological by nature and interdisciplinary in scope. It involves investigation, analysis, synthesis and presentation of information in electronic form. It studies how these media affect the disciplines in which they are used, and what these disciplines have to contribute to our knowledge of computing.

For Kirschenbaum this definition is at once sufficiently accurate, but also, importantly, capacious. His reasons that English is a good home for research of this kind include: that text is a tractable medium for computational analysis; that there have been many conversations around critical editing and the use of computers since the 1990s; that there is a long history of the use of computers in writing; that there is a convergence between teaching composition and the use of computing; that English departments are open to cultural studies and so digital cultural artefacts can be regarded as valid subjects for study; and that there has been an explosion of interest in e-reading. By contrast, Stephen Ramsay, also a scholar of English, at the MLA conference in 2011 claimed that to be a digital humanist you have to be building something, and that you need to know how to code. 'Building is,' he says, 'for us, a new kind of hermeneutic—one that is quite a bit more radical than taking the traditional methods of humanistic inquiry and applying them to digital objects' (Ramsay, 2011). But what do these definitions and demands mean for most scholars in literary subjects?

English may seem now like a natural home (to some, and generally from a US perspective), but in fact classics and religious studies were among the earliest humanities disciplines to embrace what was then called humanities computing. Father Roberto Busa's famous challenge to Thomas Watson of IBM to help him analyse the works of Thomas Aquinas was in 1949—only four years after the first stored program computer was developed. Busa realized immediately that something that could manipulate numbers could also manipulate letters, and asked Watson to sponsor the *Index Thomisticus* (Busa, 1980).

Given that computers are machines for manipulating symbols, and language is a symbolic structure, it is no accident that linguistics and the linguistic disciplines in the humanities found it more comfortable to adopt these new methods than other humanities disciplines. However, 'comfortable' is a relative term. Even now, the serious use of digital methods requires training, support and often funding. In the early days, computational manipulation was performed using a mainframe computer which lived in the university computing services with terminals scattered around campus. Text entry was initially achieved using punched cards or tape; later it could be entered via the terminal or eventually a PC. Initially, scholars often devised their own entry and processing codes, which resulted in chaos if a scholar tried to reuse texts produced in an arcane code. Indeed, it was often faster to start again than to work with texts marked up in non-standard forms, and therefore standardized coding systems were developed, based on XML (eXtensible Markup Language). As well as learning text coding, scholars had to learn computer programming and some basic statistics to perform textual analysis. One scholar (the philosopher Anthony Kenny) learnt the statistics so thoroughly that he even wrote a book to teach the basics of statistics to other humanists (Kenny, 1980). Despite the cumbersome requirements, and the extremity of the barrier to entry for many scholars, however, markup languages have proved vital for producing searchable corpora, dramatically changing the landscape of who can, and what it means to, do humanities research. Searchability has enabled the teasing out of hidden links between texts, word-frequency analyses, and, more prosaically, made scholarship from home, or abroad, viable or even preferable. Early problems with technicality, that still persist today if in (somewhat) less dramatic form, were always outweighed by the potential for dissemination and reflection for those who could see it, an impulse which might be traced back to Vannevar Bush's 1945 proto-hypertext system, the Memex, described in an article for *Atlantic Monthly* called 'As we may think'. Digital theorists of the

early nineties in particular, many working in areas which would later come under the umbrella of the Digital Humanities, often referred to Bush's work and in describing the Memex's augmentation of memory and recall Bush was at pains to emphasize his hypothetical machine's inbuilt relationship with a system that was already present:

> The human mind operates by association. With one term in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature. (Bush, p.44)

In this quotation, and others like it drawn from similar work by early digital pioneers (see for example the work of Theodore (Ted) Nelson), it was frequently argued that one could see how the linking made possible by markup languages might make manifest the asymmetrical linkages that had been in our cultural products all along, connections which were in turn relevant both to how the mind worked and to the arguments for intricate but non-hierarchical relationships that were posited as tenets of the linguistic turn, a postmodern outlook, or many of the philosophies and methodologies that they inspired. J. David Bolter, for instance, suggested in his *Writing Space* (1991) that:

> [a]s long as the printed book remains the primary medium of literature, traditional views of the author as authority and of literature as monument will remain convincing for most readers. The electronic medium, however, threatens to bring down the whole edifice at once. It complicates our understanding of literature as either mimesis or expression, it denies the fixity of the text, and it questions the authority of the author. (p.153)

Bush's fading, intricate web of trails inheres in this idea, as do the extensive associations between each node in the database, truisms for both brain and Memex in his eyes. As the powers of the digital equipment grew, thanks to the early development of markup languages described above, their relation with the brain's workings and their apparent emulation of theorizations of textual production and consumption increased apace. These kinds of ideas could still be found, a decade after

Bolter's comments, in influential works for DH such as Anna Everett and John T. Caldwell's *New Media* (2003):

> When we understand computerized linking as a system of 'nested narrative—a narrative within a narrative', following the rhizoplane structure, it becomes analogous to Freudian free association, which [Jean-François] Lyotard interprets as 'a way of linking one sentence with another without regard for the logical, ethical, or aesthetic value of the link'. (p.6)

Over time such notions became part of the assumed theoretical landscape for digital studies, implicit, for instance, in Hubert Dreyfus' *On the Internet*: 'With a hyperlinked database, the user is encouraged to traverse a vast network of information, all of which is equally accessible and none of which is privileged' (Dreyfus, 2005, p. 10).[1] George Landow, in *Hypertext* (1992), is perhaps the theorist to most explicitly link literary poststructuralism with the field and equipment of computer science:

> Like Barthes, Foucault, and Mikhail Bakhtin, Jacques Derrida continually uses the terms *link* (*liaison*), *web* (*toile*), *network* (*réseau*), and *interwoven* (*s'y tissent*), which cry out for hypertextuality; but in contrast to Barthes who emphasizes the readerly text and its nonlinearity, Derrida emphasizes textual openness, intertextuality, and the irrelevance of distinctions between inside and outside a particular text. (p. 8)

And yet more baldly, Landow sees hypertext as 'an almost embarrassingly literal embodiment' of such theory (1992, p. 34).[2] Over the early twenty-first century, the extremity of such assertions has tended to be moderated down, but, as Marie-Laure Ryan states:

> it is easy to see how the feature of interactivity conferred upon the text by electronic technology came to be regarded as the fulfilment of the postmodern conception of meaning. Interactivity transposes the ideal of an endlessly self-renewable text from the level of the signified to the level of the signifier. (2001, p. 5)

Such interactivity saw its genesis in the range of difficult early experiments in markup.

In classics and linguistics, the building of large corpora was similarly an early computational task. The *Thesaurus Linguae Graecae*

(*www.tlg.uci.edu*) began to build its corpus of all extant Greek literature in 1971, drawing on a tradition of classical studies reaching back four centuries. Now, the TLG Digital Library contains virtually all Greek texts surviving from the period between Homer (eighth century BCE) and the fall of Byzantium in 1453 and is available online for a modest subscription. In the English language the definition of what is a 'large' corpus has changed and grown as power, storage and connectivity have increased exponentially. At the end of the 1960s, the Brown Corpus of American English was developed containing samples from around 500 texts, totalling around one million words. In the early 1990s work began on the British National Corpus (*www.natcorp.ox.ac.uk*), now a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the twentieth century, both spoken and written. The Corpus of Contemporary American English (COCA) with 450 million words claims to be the largest freely available corpus of English, and the only large and balanced corpus of American English (http://corpus.byu.edu/coca). Once large-scale corpora are available, they can be used and analysed in many different ways: patterns of word usage across time and geographic regions can be tracked; dialectal variations can be mapped; language usage from different population groups can be compared: children; women/men; different Englishes compared (UK/US/Australian/Creoles etc.).

Early computational work in English was largely mathematical and statistical, dictated by what the computer could do best. So computational stylistics and authorship studies dominated, as a glance through the tables of contents of the journal *Literary and Linguistic Computing* (now known as *Digital Scholarship in the Humanities*) from the 1980s will confirm. This had minimal effect on the mainstream of literary criticism, which was dominated by more theoretical modes of enquiry: structuralism and post-structuralism, Marxism, feminism etc. In the 1990s, this began to change as computers were better able to handle non-textual materials and hypertextual modes of representation began to be possible. These changes were largely brought about by developments in the technology such as Douglas Englebart's work on graphical user interfaces and his invention of the mouse, and Apple's development of the Macintosh computer. These new technical developments brought a whole new dimension to the use of computing in the humanities: visual images, and eventually sound and motion, could be incorporated alongside textual materials and linked and navigated with ease. In 1980, Ted Nelson published his seminal work

*Literary Machines* where he coined the term 'hypertext' and imagined many of the hypertextual features that we take for granted today. He has been playing with these ideas since the 1960s (see e.g. Nelson 1965), but lacked the computational facilities to make them a reality.

A seminal moment in literary computing was Apple's development of the hypertext program Hypercard in 1987, with the wonderful promotional statement, referencing Bush again, 'The human mind works by association, so why don't computers?' This put a pre-World Wide Web hypertext creation package in the hands of anyone with a Macintosh computer, and indeed had more functionality and flexibility than the Web was to have for many years. Other hypertext authoring programmes with graphical user interfaces also appeared around that time, available across both Macintosh and PC computers. These programmes were easy for end-users to master, and literary hypertexts exploded, moving literary computing away from the numerical and into the exploratory and analytical. These hypertexts included Patrick Conner's *Beowulf Workstation* for teaching the Anglo-Saxon poem; Michael Best's *Shakespeare's Life and Times* CD-ROM, now developed as an internet resource and available at *http://internetshakespeare.uvic.ca*; and *CD Word,* the first digital library of serious Bible study tools with commentaries, lexicons, and support for Greek, Hebrew, and extensive hypertext linking. In 1989, George Landow developed *The Dickens Web* in Intermedia, and he also began to develop the theoretic formulation for literary hypertexts described above. *The Dickens Web* situated *Great Expectations* in a complex network of contexts and relationship, dealing with Dickens' life and literary connections, as well as related subjects such as Victorian History, history of public health, religion etc. *The Dickens Web* is still available from Eastgate Systems (*www.eastgate.com/catalog/Dickens.html*). Hypertext and hypermedia systems proved to be ideal platforms for creating learning materials for students. The English Department at the University of Glasgow was an early adopter of such tools, and created a suite of learning materials known as STELLA (Software for Teaching English Language and Literature and its Assessment). The STELLA materials have been in existence for over 20 years now, have been successfully migrated across generations of technology, and have taught thousands of students the basics of Old English, and English and Scottish Linguistics (www.gla.ac.uk/schools/critical/aboutus/resources/stella/).

The biggest problem for the literary hypertexts being developed at the end of the 1980s and the beginning of the 1990s was the rapid pace of change of the software and operating systems. CD-Word, cited

above, was developed at huge cost under Windows 2.0 and delivered on a CD. When the operating system changed, it became too costly to keep up the development and support. When the Web came along at the beginning of the 1990s, the death knell was sounded for many of these systems. The Web, fundamentally based on hypertext and interlinking, was a triumph initially of connectivity and standardization over function. Early web hypertexts seemed much more primitive than their locally based predecessors, but this was soon to change; many of the literary hypertexts mentioned above have since migrated, changed and flourished in a Web environment.

Another critical early problem with digital literary texts was the lack of standardization of encoding. Texts needed to be translated into forms that the computer could understand, initially using a limited range of ASCII character codes. Even with a more flexible range of possibilities, for texts to be processed, exchanged and analysed, standard forms of encoding needed to be developed. The Text Encoding Initiative consortium, over more than 25 years, has been working to develop and maintain a standard markup for the representation of texts in digital form. Early text processing took a presentational view of markup: for instance, an element was described as italic without defining *why* it was italic. The TEI guidelines (2007), using XML, define markup structurally, that is, they describe an element by its function—heading, emphasis, foreign word in the text—with its presentation left to a later rendering.

At the end of the 1980s, scholarly editors began to think of the use of computer tools not just as means of producing printed texts, but as means of displaying editions electronically. For what is an edition if it is not a hypertext, a complex web-like system of linked transcriptions, variants, glosses, notes, and all the other apparata we associate with the printed form? Textual critics like Jerome McGann, Kathryn Sutherland, Peter Shillingsburg, Peter Robinson and others began to both theorize about literary works and editions as hypertexts and to develop hypertext and multimedia systems around these ideas. Particularly influential were Robinson's work on Chaucer's *Canterbury Tales* and McGann's edition of the works of Dante Gabriel Rossetti (of which, more below). Initially, ideas (as they do) ran faster than practical possibilities. The fluidity of the electronic medium was seen as a benefit: how wonderful to be able to spot an error and correct it instantly, to be able to add a reference, to constantly add new transcriptions or images to online editions, to interlink internal and external referents and create complex paths through the materials. Peter Robinson proposes a new model of editing where the edition is made by the reader from whatever is available,

the reader determines what is read and how it is presented, the reader controls the choice of materials and anyone can alter any word and invite others to read the altered text. He calls this model 'fluid, collaborative and distributed editions' (Robinson, 2009, paragraph 33). However, this brings up a crucial issue: developing complex functions makes editions much harder to preserve for the long term. Another feature that editions *must* have is durability, something print has of course always achieved. Take for example R.W. Chapman's 1923 edition of the *Works of Jane Austen* published by the Clarendon Press. In 1966, a revised edition was published, which was reissued in 2001. The text set in 1923 was used in both later editions, so the page numbers never changed, citation was utterly stable, and even the 1923 text is as readable today as the day it was published. These issues were kept very much in mind by the team that developed the *Jane Austen's Fiction Manuscripts Digital Edition* edited by Kathryn Sutherland (**www.janeausten.ac.uk**), which brings together around 1100 pages of fiction written in Jane Austen's own hand. Through digital reunification, it is now possible to access, read, and compare high quality images of original manuscripts whose material forms are scattered around the world in libraries and private collections. Also provided are newly produced transcriptions that can be accessed alongside the manuscripts, and detailed headnotes for each manuscript. This is one of the earliest collections of creative writings in the author's hand to survive for a British novelist. An earlier example of the power of the computer to unite and display literary materials linked to their physical manifestations is the Rossetti Archive, begun in 1993 by Jerome McGann and completed in 2008. Rossetti was a painter, designer, writer, and translator, and so his works were produced across many different media, which meant that the print medium did not necessarily serve them well; the digital medium is ideal for presentation and cross-linking of such a complex oeuvre (www.rossettiarchive.org).

An interesting question that arises when editions are produced digitally, with the inclusion of so much source and explanatory material, is, what *is* an edition and when does it become an archive? In the print world, an edition is finished, published and used, often for decades. Critical comments on the edition can be made, but they are always outside the work itself. In the digital world, everything *about* the edition can become *part of* the edition, adding always to this growing archive of materials. And this also begs the question of the role of the editor, which may be different from that in the print world. In multi-partner digital projects, there is a whole range of roles and responsibilities, some of which map onto those functions in the print world, while some

don't—managing the complex technical dimensions of the edition, for example. Multi-partner editing projects are not just the province of the digital, of course: there are many, complex print-based editing endeavours and there are some interesting hybrids. The *Complete Works of Jonathan Swift* (2008) have been published in an 18-volume print edition by Cambridge University Press. These are accompanied by a freely accessible electronic archive containing around 300 texts, including documentary transcriptions of Swift's works as they appear in their original printed editions (generally, first editions), as well as other materials. The intention here is to marry what print does best with what digital does best, producing both a scholarly edition and an archive. Cambridge have also published the complete works of Ben Jonson (2008) in print, accompanied by an online scholarly digital edition, incorporating old-spelling texts and digital images of manuscripts and major early editions.

It is not just scholars themselves who have been making available texts and sources; there has also been intense activity in libraries and in the commercial world, sometimes in partnership. Early English Books Online (EEBO, http://eebo.chadwyck.com/) and Eighteenth-Century Collections Online (ECCO, http://gdc.gale.com/products/eighteenth-century-collections-online/) have put millions of pages from hundreds of thousands of books onto the desktops of scholars, with catalogue records and in most cases full text searchability. Google Books also offers up millions of volumes from libraries all over the world; though often of patchy quality and with little metadata, they are still an enormous boon, especially for those working outside of the developed world. Probably the most important online initiative for the study of English being undertaken by publishers is the Oxford University Press initiative Oxford Scholarly Editions Online (OSEO, www.oxfordscholarlyeditions.com). This started as a republishing in complex digital form of editions published by the Press itself, and has developed into much more extensive coverage of scholarly texts by licensing works from other major publishers. Currently, OSEO provides access to more than 450 scholarly editions of material written between 1485 and 1788, including all of Shakespeare's plays, the poetry of John Donne, and works by John Milton and John Locke. These editions contain over 44,000 different works including more than 400 plays, over 17,000 poems, and more than 26,000 other works, the equivalent of over 233,000 print pages. All works are rekeyed and tagged with XML markup, and are presented using advanced searching and linking abilities, with accompanying PDFs of the original print page. Some difficult decisions had to be made by the OSEO editorial team and these were the subject of much debate

and discussion. For instance, the decision was taken to reprocess the original print editions for online presentation, without re-editing or updating them. The reason for this was total fidelity to the original publication, even to the point of reproducing known errors. Where possible these errors are signalled, but they are not corrected. PDFs of the original editions are available so that users can always check a reading in a faithful representation of the original. As a major resource like this grows over many years there will be more hard decisions to debate. For example, how can other online materials be linked in that may reside behind other paywalls?

As things get ever bigger and more extensive, new paradigms present themselves: Big Data and Distant Reading, to name but two. Big Data has been claimed as the Next Big Thing in the humanities, and is already seen as the current Big Thing in other disciplines. Big Data is the term used for collections of data that are orders of magnitude larger than the corpora such as the Brown Corpus or the British National Corpus discussed above. Big Data is not necessarily created for a particular purpose and marked up in a systematic way: it can be derived from many different sources and may not be standardized, so may need different approaches and new tools. Many of the commercial tools for processing large-scale data grow out of the military and surveillance communities, and in order to make sense of the results of data processing, data visualization methods have been developed. There is some discomfort among humanists in regarding the objects of our study as 'data', feeling that this is something of a reductive term. But, for scholars of English, there are new opportunities for research opened up by the massive availability of text online, especially if these are published as open data, freely accessible. Just to give one example, hundreds of millions of pages of historic newspapers have been digitized over the last ten years by newspapers themselves, by libraries, and by third-party commercial publishers like NewspaperARCHIVE.com, which calls itself the world's largest collection. To illustrate the possibilities, a search in Papers Past (paperspast.natlib.govt.nz/), the New Zealand newspaper archive, for 'Charles Dickens' yielded 16,654 results. An early one chosen at random led to an 1873 review of John Forster's life of Dickens in the *Wellington Independent*. Imagine the new research possible into, for instance, publication and reception throughout the English-speaking world. And imagine aggregating such searches across library catalogues, dictionaries, biographical dictionaries, letters as well as newspapers. That is what Franco Moretti has termed 'distant reading', a different view of our textual universe, taken from afar, which can pinpoint new

lines of enquiry and reveal new constellations of relationships. Moretti is a leading scholar of European literature who founded the Stanford Literary Lab and who uses technical methods to survey vast swathes of literary works: turning literature into data in order to identify patterns that are difficult if not impossible to see with traditional approaches. Moretti coined the term in a 2000 article in the *New Left Review* and it has entered the lexicon of digital humanities, meaning different things to different scholars. For Moretti, it constituted a new science 'where a new problem is pursued by a new method' (Moretti, 2000 and 2013, p. 55). The present authors prefer to think of it as a set of new methods to pursue many paths of enquiry, old and new, and as an adjunct to our traditional methods of work, not a replacement for them.

## The future of the book

From very early in the development of digital humanities, and particularly in departments of English, a great deal of discussion has centred on the future of the print medium in the digital age. The publishing industry has always been ready to embrace new technologies in the pursuit of better, cheaper and faster production of its wares, and scholars and writers saw advantages in the new media for the development and promulgation of literary forms. Digital technology has been embedded in all forms of print production for decades, and it has been a logical progression to digital access, especially in journal provision. At the end of the 1980s, doomsayers began predicting the end of the printed book, but as a form it is proving surprisingly robust. Early predictions posited that CD-ROM-based and then web-based hypertext and multimedia literary forms would proliferate and the linear printed book would die away. Interestingly, this has not happened and does not seem likely to happen in the near future. What *has* happened is a huge rise in the popularity of ebooks, which, on the whole, mimic almost exactly the print form, with some added functionality such as the ability to search, annotate, link out to dictionaries, etc.

Pundits tend to predict that a new technology is likely to supersede previous ones. Some do: the telephone killed off the telegraph, for example; and some don't: television was supposed to kill off radio, but radio survived, probably because it is possible (and often desirable) to listen to the radio while doing other things like driving or ironing. The problem with trying to discuss print books versus ebooks is that, for many, the opposition is printed word equals linear and static, digital word equals non-linear and dynamic; printed word is fixed, digital

is interactive, and these are of course false notions. Very few printed products are strictly linear; even novels and poetry, though apparently linear, play with and rupture linearity in interesting ways, but periodical publications (newspapers, magazines, journals) also defy (and resist) linearity, as do reference works and complex works of scholarship. And every text is 'interactive', changing according to a particular reader at a particular hour in a particular place. All readers create their own text while reading, and every new reading is an act of recreating, just as every new access to an electronic text is a process of creating a human-readable version afresh on the screen.

An interesting development that has happened alongside digital advances is a degree of creativity by authors and designers in the development of the print book, resulting in books that cannot properly be represented in digital form. For instance, Jonathan Safran Foer's *Tree of Codes*, published by Visual Editions, is a book crafted out of another book: *The Street of Crocodiles* by Bruno Schulz, a collection of short stories published in Polish in 1934 and translated into English in 1963. *Tree of Codes* is a novel, a text and an art object and is resolutely physical in format: the pages of the Schultz collection are sculpted into the new work, with words physically cut from the pages to reveal the Foer work. *S*, a 2013 mystery novel by J. J. Abrams and Doug Dorst, is described by Joshua Rothman in the *New Yorker* as 'the best-looking book I've ever seen'. Impossible to produce or reproduce in other than book form, this work looks like an old library book called *Ship of Theseus*, which forms the central text of the work. But around this text is another text, written in the margins, in inserts of postcards, photographs, even a map. *S* was described by another reviewer as 'a celebration of the book as a physical thing, possessor of wonders that cannot be translated into digital bits' (Tsouderos, 2013). Anyone who thinks the printed book is dead is invited to contemplate the lengths to which authors and designers can go to prove them wrong.

## What to leave out

English researchers all use some aspect of digital technology in their research; at the very least no one is safe from Google and word processing. But as the wealth of tools and methods available to academics proliferate, and their prioritization by funding bodies increases, so too does an interesting question: 'am I doing DH?' As one colleague put it: 'What is it with the digital humanities? No one I work with seems to be asking "am I doing manuscript studies?!"' For the foreseeable future this

is a question that will, and maybe must (or should) haunt the digital humanities. But for those who remain sceptical of DH's place in the English department there's also a second worry: what if we've accidentally been doing it all along?

Such queries arise despite the relative vintage of DH, presumably as an effect of the instability of the underlying architecture of the research. Has there ever before been an object that has so profoundly influenced Humanities study whilst changing at the pace of Moore's Law, having the capacity to become exponentially more complex, more powerful every 18 months? Above, we've outlined a range of approaches to researching in English with the addition of digital tools, many of which will be recognizable in some form or another for most Humanities researchers. And yet, despite the identification of distinctive 'waves' across the years, rises and falls of what's important, what's possible, what's desirable, even this incomplete and selective collection demonstrates what can feel like a baffling lack of coherency in what is meant to be a discipline, as DH is now assumed to be. If DH is indeed a distinctive field, what is it for, what are its goals? These are important questions, but ones that have often been sacrificed for the converse, slightly easier, less satisfying approach: what *isn't* it, what can we safely ignore?

The theme of the 2011 Digital Humanities conference at Stanford was 'Big Tent Digital Humanities'. Positioned as a direct response to the increasing diversity of practices that were being labelled as 'Digital Humanities', the conference aimed to realize fully the debate about the boundaries of the discipline, to question whether there was a 'right' way to do DH and, implicitly, whether there was a wrong way. Maybe 'wrong' is the wrong word, but (as now) there was a legitimate concern amongst some practitioners of digital humanities research that the term was being spread too thin, or, worse, reduced to a buzzword, excusing the same old scholarship just because it had a website and Twitter feed written into the grant proposal. Such concerns needed to be addressed (and perhaps put aside), and the conference aimed to celebrate the 'big tent', the inclusion of the greatest diversity of responsible research. This didn't stop the on-going discussion of exactly what DH is and the continued deployment of a slogan that had quickly slipped into parody in some circles: 'more hack, less yack.' The phrase was, in fact, always a joke, as Bethany Nowviskie noted in her charting of its origin (Nowviskie, 2014), but it still came to stand in for a sense that 'true' DH was research which *did*, which produced, which built things. What Nowviskie rightly critiques in that origin story, however, is the

sterile split that the phrase implies between a healthy practice and an unhealthy mere theorizing:

> In my view, to pretend or believe that 'more hack; less yack' represents *a fundamental opposition in thinking* between humanities theorists and deliberately anti-theoretical DH 'builders' is to ignore the specific history and different resonances of the phrase, and to fall into precisely the sort of zero-sum logic it seems to imply. Humanities disciplines and methods themselves are not either/or affairs. The humanities is both/and.

Humanities research is, indeed, a both/and set of disciplines, but that it now includes *both* reading *and* coding digital creations is a shift. It changes our practices dramatically (researchers who no longer see the book or article, but the database or app as the desirable output), subtly (the implications of accessing source texts digitally and the increasing searchability of metadata), and fundamentally (the lone researcher in the library is no longer able to accomplish all that they might want to; collaboration and multi-authored projects challenge the viability or desirability of the garret or ivory tower as a workspace). So the idea of 'more hack, less yack' remains significant, not as an accurate description, but as an indicator of the recognition of change; for anyone who has felt empowered by that phrase, even for a moment, it's been about changing what is viable.

This emphasis on building, on the carpentry of DH, returns the Humanities to an older debate about methods and what they best reveal. Building certainly isn't all of what the digital humanities is, or can be, but it puts it on the same continuum of advocacy that has seen creative writing and other art practices positioned as viable research methodologies that can reveal something distinctive. Writing a poem can tell you something different about the act of writing and of contemplation than criticism, but it is not the whole of English Studies; painting a picture can tell you about brushwork, and building a trebuchet with original tools can tell you about the knowledge required to get there, but these aren't the sum of Art History or Archaeology. Building in DH, similarly, does more than just make things (otherwise it would simply be fabrication), but it is also not all that DH can or should be.

In thinking the above ideas through it is perhaps useful to consider an edge case. At the time of writing, Matt is working with the Royal Shakespeare Company and a husband and wife team of artists, Davy and Kristin McGuire, on a project to develop a pop-up book version of

scenes from Macbeth.[3] What's unique about the McGuires' approach is that the beautifully cut and folded scenes that unfurl from each page also come alive: through a system of digital projection and reflection, cleverly hidden within the body of the book itself, characters walk across scenes that flicker with energy and movement even as they're newly minted with each turn of the page. It's captivating and always makes its audience become children again—it feels magic and intimate, like hiding under the covers with a torch and escaping the burden of sleep by heading to another world. [4]

But the McGuires aren't doing digital humanities, or at least not exclusively. They're making art and, by devising the mechanisms by which their art functions, engaging in design, prototyping, and fabrication, reminiscent of *Tree of Codes* and *S* mentioned above. These are elements of digital humanities work, but they are not its sum or sole components. The Royal Shakespeare Company aren't doing digital humanities; they're providing vital support for the project, acting as producers and putting the other participants in contact with Shakespeare and his plays and coding and production experts. And Matt isn't doing digital humanities; he's writing a series of essays to accompany the project, like extended variations on the gallery blurb beside a painting, to help an audience, the RSC, maybe even the McGuires themselves, to articulate what's going on, why and how the object and the experience feels so rich, overdetermined, why it means so much. The project, however, seems to be absolutely a DH project—there's something in the sum, and this suggests that that question of 'am I doing DH?' remains uneasy if it's neatly reduced to practices or outcomes.

There is something genuinely distinctive about digitization and its effects that requires a new way of working and speaking, a new set of sensitivities, and it is here, perhaps, that we might like to identify the digital humanities. Actually, maybe the concerns aren't so new, but rather a reconfiguration with the net cast a little wider. The conversations that the McGuires' work and other experimental print-based books prompt—about the continuing importance of stories on paper in a digital age, about the potential inherent in drawing on old myths and new technology, the pure and devastating drives of memory and hope—also demand what any sensitive reading of a text has always required: paying attention to the conditions of its production and reception. Collaborations like the above give each member new ways of considering the objects under discussion, but despite the importance of building for DH, as both source and provocation, it also requires a significant critical component so that it doesn't have politics removed

from its concerns. This, in turn, demands that researchers have at least a basic knowledge of and interest in the popular technologies of content access, the practices of using those technologies, and the kinds of cultural forces that surround them, what we might call a cyberculture or digital culture. To not be attentive to these wider concerns around digital technologies would be like studying Victorian novels without considering Empire, industry, or urban sprawl; it would be like studying Shakespeare's manuscripts without thinking of where the plays were performed and how they were received. At this stage, then, DH in English departments can't be just reading literature with new methods, *or* making new things, *or* talking about new things in new ways—it must be 'both/and', it must be 'all'.

The tent, then, isn't big; it's vast, drawing on everything available to ask what makes the digital distinctive as an object or method or product of study. We might find that what this means is that we can all do DH, all contribute, if we're not already, to a project that may end up increasingly delineated and coherent, with less mutable values and practices. Or, instead, we're already finding ourselves on a path to where the Digital Humanities simply become the Humanities, the same plural concerns and methods made a little better, a little richer, and more able to deal with the realities of increasingly ubiquitous computing.

## Notes

1. For a snapshot of the mid-nineties critical theoretical responses to hypertext and electronic reading environments which laid the groundwork for Dreyfus' pronouncement see the Landow edited collection *Hyper / Text / Theory*.
2. Matt Hayler further explores this history of metaphors inherent in the digital in an unpublished section of his doctoral thesis (Hayler, 2011).
3. The project is funded by REACT, an AHRC knowledge-exchange hub that aims to unite academic researchers with non-academic research partners.
4. For an example of the kinds of effect, though achieved by a slightly different method of projection, see their *The Ice Book* project (www.theicebook.com).

## References

Bolter, J. D. (1991) *Writing Space: The Computer, Hypertext, and the History of Writing* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Busa, R. (1980) 'The annals of humanities computing: the *Index Thomisticus*', *Computers and the Humanities*, 14, 83–90.

Bush, V. (1945) 'As we may think', *Atlantic Monthly*, 176, 101–108. Rpt. in *The New Media Reader* (2003) Wardrip-Fruin, N. and Montfort, N. (eds), (Cambridge, MA: MIT Press), 37–47. A simulation of Bush's design can now be found online at http://memexsim.sourceforge.net.

*Cambridge Edition of the Works of Ben Jonson*, 7 vols (2012), (Cambridge: Cambridge University Press), http://universitypublishingonline.org/cambridge/benjonson/.

*Cambridge Works of Jonathan Swift*, 18 vols (2008), (Cambridge: Cambridge University Press), http://jonathanswiftarchive.org.uk/index.html.

Dreyfus, H. (2005) *On the Internet* (London: Routledge).

Everett, A. and Caldwell, J. T. (eds) (2003) *New Media: Theories and Practices of Digitextuality* (New York: Routledge).

Foer, J. S. (2010) *Tree of Codes* (London: Visual Editions), http://www.visual-editions.com/our-books/tree-of-codes.

Hayler, M. S. (2011) 'Incorporating Technology: A Phenomenological Approach to the Study of Artefacts and the Popular Resistance to E-reading', PhD thesis, University of Exeter, https://ore.exeter.ac.uk/repository/handle/10036/3615.

Kenny, A. (1980) *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities* (Oxford & New York: Pergamon Press).

Kirschenbaum, M. G. (2010) 'What is digital humanities and what's it doing in English departments?', *ADE Bulletin*, 150, http://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf.

Landow, G. P. (1992) *Hypertext: The Convergence of Contemporary Critical Theory and Technology* (Baltimore: Johns Hopkins UP).

Moretti, F. (2000) 'Conjectures on world literature', *New Left Review,* 1, http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature.

Moretti, F. (2013) *Distant Reading* (London: Verso).

Nelson, T. H. (1965) 'A file structure for the complex, the changing, and the indeterminate', *Association for Computing Machinery: Proceedings of the 20th National Conference,* 20, 84–100. Rpt. in *The New Media Reader* (2003) Wardrip-Fruin, N. and Montfort, N. (eds), (Cambridge, MA: MIT Press), 134–145.

Nelson, T. (1980) *Literary Machines* (Sausalito, CA: Mindful Press).

Nowviskie, B. (2014) 'On the origin of "Hack" and "Yack"', http://nowviskie.org/2014/on-the-origin-of-hack-and-yack/.

Parker, J. (2012) 'Editorial: Digital humanities, digital futures', *Arts & Humanities in Higher Education,* 11 (1–2), 3–7.

Ramsay, S. (2011) 'On building', http://stephenramsay.us/text/2011/01/11/on-building/.

Robinson, P. M. W. (2009) 'The ends of editing', *Digital Humanities Quarterly*, 3 (3), http://digitalhumanities.org/dhq/vol/3/3/000051/000051.html.

Rothman, J. (2013) 'The story of "S": Talking with J. J. Abrams and Doug Dorst', *The New Yorker*, 23 November, http://www.newyorker.com/books/page-turner/the-story-of-s-talking-with-j-j-abrams-and-doug-dorst.

Ryan, M.-L. (2001) *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media* (Baltimore: Johns Hopkins UP).

TEI *Guidelines for Electronic Text Encoding and Interchange* (2007) http://www.tei-c.org/Guidelines/.

Tsouderos, T. (2013) '*S.* by J. J. Abrams and Doug Dorst' (review), *Chicago Tribune,* 28 November, http://articles.chicagotribune.com/2013-11-28/features/chi-jj-abrams-s-review-20131128_1_j-j-abrams-printers-row-journal-theseus.