

Print Scholarship and Digital Resources¹

Whosoever loves not picture, is injurious to truth: and all the wisdom of poetry. Picture is the invention of heaven: the most ancient, and most akin to nature. It is itself a silent work: and always of one and the same habit: yet it doth so enter, and penetrate the inmost affection (being done by an excellent artificer) as sometimes it o'ercomes the power of speech and oratory.

Ben Johnson, *Explorata or Discoveries*, ll. 1882-90.

Introduction

In the late 1990s there was a great deal of concern about the death of the book. From every corner it was possible to hear quoted Victor Hugo's Archbishop, complaining that 'ceci, tuera cela' (Nunberg, 1996). Articles and books were published on the future of the book, which was assumed to be going to be a brief one. (Finneran, 1996). At the same time we were being told that we might no longer need to commute to work, or attend or teach at real physical universities, and of course if there were no longer any books, we would only need virtual libraries from which to access our electronic documents. Just a few years later all this seems to be as misguidedly futuristic as those 1970s newspaper articles predicting that by the year 2000 we would all eat protein pills instead of food. It is clear then than far from being killed off, print scholarship is still very much alive and well, and that its relationship to electronic resources is a highly complex one. In this chapter I will examine this relationship, and argue that we cannot hope to

understand the complexity of such a relationship without looking at scholarly practices, and the way that such resources are used. This in turn means examining how we use information, either computationally or in print, in the wider context of scholarly life. We must consider how we read texts, and indeed visual objects, to understand why print remains important, and how it may relate to the digital objects and surrogates that are its companions.

Scenes from academic life

To exemplify some of the changes that are taking place I would like to recreate two scenes from my academic life, which are illustrative of some of these larger issues.

Thessaloniki, Greece:

I am in the Byzantine museum, being shown some of the treasures of this northern Greek city, which prides itself on having been longest under continuous Byzantine rule. It is full of icons, mosaics, marble carvings and richly painted tombs. My guide asks what I think of them, and repeatedly I marvel at their beauty. But this, it appears, is not the point. We must, I am told, be able to read these images and icons in the way that their creators intended us to. I see a tomb with a painted tree sheltering a rotund bird which looks rather like a turkey. In fact this is an olive, signifying eternity and peace, and the bird is a peacock, symbol of paradise. This is not simply wallpaper for the dead, but a statement of belief. The icons themselves glow with colours, whose richness I wonder at, but again they must be

read. The gold background symbolises heaven, and eternity, the red of a cloak is for love, the green band around the Madonna's head is for hope. The wrinkles painted on her face are to symbolise that beauty comes from within, and is not altered by age. My guide begins to test me, what do I see in front of me? I struggle to remember the unfamiliar visual alphabet that I am being taught, and realise that reliance on the printed word, and a lack of such meaningful images, has deprived me of a visual vocabulary; that such iconography is part of another tradition of communication, whose roots, at least in this part of Greece, are extremely ancient. They have long coexisted with the culture of the printed word, but have not entirely been supplanted by it.

Portsmouth, England:

I am on an advisory board for the Portsmouth record office. Over a period of several decades they have painstakingly produced nine edited volumes of printed work cataloguing some of holdings of their archives. They are arranged thematically, concerning dockyards, houses in the old town, legal documents. All are handsome hardbacks with an individual design. There is, it seems, still a substantial backlog of volumes in preparation, but printing has become so costly that they have decided to publish electronically, which is why I am here. We spend time discussing volumes awaiting release, and all are relieved to find that the next volume should soon be ready after a period of thirty years in preparation.

There is a certain culture shock on both sides. I am amazed to discover how long the process of editing and publication takes. Thirty years is long, but an average of

a decade seems to be quite usual. I reflect that the papers themselves are historic, and still exist in the archive, waiting to be discovered, even if not calendared. But from the world I am used to, where technology changes so quickly, it is hard to return to a sense of such relative lack of change and urgency.

A fellow panel member suggests that what had been thought of as several discreet print volumes, on large scale maps, title deeds, city plans and population data could easily be combined, in an electronic environment, with the help of GIS technology. I in turn argue that the idea of separate volumes need not be retained in an electronic publication. There is no need to draw up a publication schedule as has been done with the print volumes. We could publish data on different themes concurrently, in several releases, when it is ready, so that the digital records will grow at the pace of those who are editing, and not have to suffer delays.

These suggestions are welcomed, enthusiastically but the series editor reminds us that there are human dimensions to this process. Those editing individual volumes may want to see their work identified as a discrete entity, to gain credit from funding authorities or promotion boards. These collections would also, it seems, be incomplete without an introduction, and that, paradoxically is usually written last: a major intellectual task and perhaps the factor, I speculate privately, which may have delayed publication, since it involves the synthesis of a vast range of sources and an attempt to indicate their intellectual value to a historian. Would it be possible to publish a release of the data without such an introduction, even if temporarily? We explore different possibilities for ways that data might appear with different views, to accommodate such problems, and the intellectual

adjustments on both sides are almost visible. The historians and archivists contemplate the electronic unbinding of the volume as a controlling entity: those interested in computing are reminded that more traditional elements of the intellectual culture we are working in cannot be ignored of the project is to keep the good will of the scholars who work on it.

Tools to think with

These two examples serve as illustration of some of the very complex issues that we must contend with when considering the relationship between printed and electronic resources. As Jerome McGann argues, we have grown used to books as the primary tools to think with in the humanities. We may read one text, but are likely to use a variety of other books as tools, as we attempt to interpret texts of all kinds. (McGann, 2001, Chapter 2). In the remainder of this chapter I shall demonstrate that what links the two examples that I have quoted above with McGann's concerns is the question of how we use materials in humanities scholarship. I shall argue that whatever format materials are in, computational methods must make us reconsider how we read. Since we are so used to the idea of reading as an interpretive strategy we risk taking it for granted, and considering it mundane when compared to exciting new computational methods. But, as the example of the Macedonian museum shows, reading is a much more broad ranging process than the comprehension of a printed text, and that this comprehension itself is a complex process which requires more detailed analysis.

It is also arguable that the visual elements of the Graphical User Interface have also made us rediscover the visual aspects of reading and comprehension. Both of these processes must be considered in order to try to understand the complex linkage between digital resources and print scholarship. Because if we assume that reading printed text is a simple process, easily replaced by computational methods of interpreting digital resources, we risk underestimating the richness and complexity of more traditional research in the humanities.

When the use of digital resources was first becoming widespread, assumptions were made that such resources could and indeed should replace the culture of interpreting printed resources by reading them. Enthusiasts championed the use of digital resources, and decried those who did not use them as ill informed or neoluddite. During the 1990s efforts were made to educate academics in the use of digital resources. Universities set up learning media units to help with the production of resources, and offered some technical support to academics, though at least in the British system this remains inadequate. The quality and quantity of digital resources available in the humanities also increased. And yet print scholarship is far from dead. Academics in the humanities still insisted on reading books, and writing articles, even if they also used or created digital resources. As I discovered in Portsmouth, cultural factors within academia are slow to change. The authority of print publication is still undoubted. Why, after all, is this collection published in book form? Books are not only convenient, but carry weight with promotion committees, funding councils and one's peers. Computational techniques, however, continue to improve and academic culture changes, even if slowly. What is not likely to change in the complex dynamics

between the two media is the fundamentals of how humanities academics work, and the way that they understand their material. How, then, should we explain the survival of reading printed texts?

We might begin this process by examining early attempts to effect such a culture change. The 1993 *Computers and the Humanities* was very much in proselytising mode. In the keynote article, of a special issue on computers and literary criticism Olsen (CHUM, 1993a) argued that scholars were being wrong headed. If only they realised what computers really are useful for, he suggested, there would be nothing to stop them using computer methodology to produce important and far-reaching literary research. This is followed by an interesting collection of articles, making stimulating methodological suggestions. All of them proceeded from the assumption that critics ought to use, and what is more, should want to use digital resources in their research. Suggestions include the use of corpora for studying intertextuality, (Greco, 1993) or cultural and social phenomena (Olsen, 1993); scientific or quantitative methodologies (Goldfield, 1993) such as those from cognitive science, (Henry and Spolsky, 1993) and Artificial Intelligence theory (Matsuba, 1993). All of these might have proved fruitful, but no work to be found in subsequent mainstream journals suggests that any literary critics took note of them.

The reason appears to be that the authors of these papers assumed that a lack of knowledge on the part of their more traditional colleagues must be causing their apparent conservatism. It appears that they did not countenance the idea that none of these suggested methods might be fit for what a critic might want to do. As

Fortier argues in one of the other articles in the volume, the true core activity of literary study is the study of the text itself, not theory, nor anything else. He suggests that: 'this is not some reactionary perversity on the part of an entire profession, but a recognition that once literature studies cease to focus on literature, they become something else: sociology, anthropology, history of culture, philosophy speculation, or what have you.' (Fortier, 1993:376) In other words these writers are offering useful suggestions about what their literary colleagues *might* do, instead of listening to critics like Fortier who are quite clear about what it is they *want* to do. Users have been introduced to all sorts of interesting things that *can* be done with computer analysis or electronic resources, but very few of them have been asked what it is that they do, and want to keep doing, which is to study texts by reading them.

As a result, there is a danger that humanities computing enthusiasts may be seen by their more traditional colleagues as wild eyed technocrats who play with computers and digital resources because they can. We may be seen as playing with technological toys, while our colleagues perform difficult interpretive tasks by reading texts without the aid of technology. So if reading is still so highly valued and widely practiced perhaps in order to be taken seriously as scholars, we as humanities computing practitioners should take the activity of reading seriously as well. As the example of my Portsmouth meeting shows, both sides of the debate will tend to make certain assumptions about scholarly practice, and it is only when we all understand and value such assumptions that we can make progress together. If reading a text is an activity that is not easily abandoned, even after academics know about and actively use digital resources, then it is important

for us to ask what then reading might be, and what kind of materials are being read. I shall consider the second part of the question first, because before we can understand analytical methods we need to look at the material under analysis.

Texts and Tradition

Norman (1999) argues that we must be aware what computer systems are good for and where they fit in with human expertise. He thinks that computers are of most use when they compliment what humans can do best. He attributes the failure or lack of use of some apparently good computer systems to the problem that they replicate what humans can do, but do it less efficiently. A clearer understanding of what humanities scholars do in their scholarship is therefore important. Computer analysis is particularly good at returning quantitative data and has most readily been adopted by scholars in fields where this kind of analysis is privileged, such as social and economic history, or linguistics. A population data set, or linguistic corpus contains data that is ideal for quantitative analysis, and other researchers can also use the same data and analysis technique to test the veracity of the results.

However, despite the pioneering work of Corns (1990) and Burrows (1987) literary text can be particularly badly suited to this type of quantitative analysis because of the kind of questions asked of the data. As Iser argues, the literary text does not describe objective reality. Literary data in particular is so complex that is not well suited to quantitative, polar opposites of being or not being, right and wrong, presence or absence, but to the analysis of subtle shades of meaning, of

what some people perceive and others do not. Its complexity is demonstrated the use of figurative language. As Van Peer (1989:303) argues this is an intrinsic feature of its 'literariness'. Thus we cannot realistically try to reduce complex texts to any sort of objective, and non-ambiguous state for fear of destroying what makes them worth reading and studying.

Computer analysis cannot 'recognise' figurative use of language. However, an electronic text might be marked up in such a way that figurative uses of words are distinguished from literal uses before any electronic analysis is embarked upon.

However there are two fundamental problems with this approach. Firstly, it is unlikely that all readers would agree on what is and is not figurative, nor on absolutely numbers of usages. As I.A. Richards' 1929 study was the first to show, readers may produce entirely different readings of the same literary text.

Secondly, the activity of performing this kind of markup would be so labour intensive that a critic might just as well read the text in the first place. Nor could it be said to be any more accurate than manual analysis, because of the uncertain nature of the data. In many ways we are still in the position that Corns complained of in 1991 when he remarked that: 'Such programmes can produce lists and tables like a medium producing ectoplasm, and what those lists and tables mean is often as mysterious.' (128). Friendlier user interfaces mean that results are easier to interpret, but the point he made is still valid. The programme can produce data, but humans are still vital in its interpretation. (Lessard and Benard, 1993)

Furthermore, interpreting the results the analysis is a particularly complex activity. One of the most fundamental ideas in the design of automatic

information retrieval systems is that the researcher must know what he or she is looking for in advance. This means that they can design the system to find this feature and that they know when they have found it, and how efficient recall is. However, unlike social scientists or linguists, humanities researchers often do not know what they are looking for before they approach a text, nor may they be immediately certain why it is significant when they find it. If computer systems are best used to find certain features, then this is problematic. They can only acquire this knowledge by reading that text, and probably many others. Otherwise they are likely to find it difficult to interpret the results of computer analysis, or indeed to know what sort of 'questions' to ask of the text in the first place.

Humanities scholars often do not need to analyse huge amounts of text to find material of interest to them. They may not need to prove a hypothesis as conclusively as possible, or build up statistical models of the occurrence of features to find them of interest, and may find the exceptional use of language or style as significant as general patterns. (Stone, 1982) They may therefore see traditional detailed reading of a relatively small amount of printed text as a more effective method of analysis.

The availability of a large variety of text types may be more important than the amount of material. Historians require a wide variety of materials such as letters, manuscripts, archival records and secondary historical sources like books and articles. (Duff and Johnson, 2002) All of these are to be found in print or manuscript sources, some of which are rare and delicate and often found only in specific archives or libraries. Of course some materials like these have been

digitised, but only a very small proportion. Even with the largest and most enthusiastic programmes of digitisation, given constraints of time, the limited budgets of libraries, and even research councils, it seems likely that this will remain the case for the foreseeable future. The historical researcher may also be looking for an unusual document whose value may not be apparent to others: the kind of document which may be ignored in selective digitisation strategies.

Indeed the experience of projects like Portsmouth records society suggests that this has already been recognised. If the actual material is so unique that a historian may need to see the actual artefact rather than a digitised surrogate, then we may be better advised to digitise catalogues, calendars and finding aids, and allow users to make use of them to access the material itself. This also returns us to the question of the visual aspect of humanities scholarship. Initially the phenomenon of a digitised surrogate increasing the demand for the original artefact seemed paradoxical to librarians and archivists. However, this acknowledges the need for us to appreciate the visual aspects of the interpretation of humanities material. A transcribed manuscript which has been digitised allows us to access the information contained in it. It may only be as a result of having seen a digital image that the scholar realises what further potential for interpretation exists, and this, it appears may only be satisfied by the artefact itself. This is such a complex process that we do not yet fully understand its significance.

The nature of the resources that humanities scholars use should begin to explain why there will continue to be a complex interaction between print and digital resources. There is not always a good fit between the needs of a humanities

scholar or the tasks that they might want to carry out, and the digital resources available. Print still fulfils many functions and this is perhaps encourages scholars to produce more, by publishing their research in printed form. But surely we might argue that computational methods would allow more powerful and subtle ways of analysing such material. Reading, after all, seems such a simple task.

What is reading?

Whatever assumptions we might be tempted to make, the activity of reading even the simplest text is a highly complex cognitive task, involving what Crowder and Wagner (1992:4) describe as ‘three stupendous achievements’ the development of spoken language, written language and literacy. Kneepkins and Zwaan (1994:126) show that:

In processing text, readers perform several basic operations. For example, they decode letters, assign meaning to words, parse the syntactic structure of the sentence, relate different words and sentences, construct a theme for the text and may infer the objectives of the author. Readers attempt to construct a coherent mental representation of the text. In this process they use their linguistic knowledge (knowledge of words, grammar), and their world knowledge (knowledge of what is possible in reality, cultural knowledge, knowledge of the theme)

These processes are necessary for even the most basic of texts, and therefore the cognitive effort necessary to process a complex text, such as a the material commonly used by humanities researchers must be correspondingly greater. Arguably the most complex of all such texts are literary ones, and thus the following section is largely concerned with such material. Although empirical studies of the reading of literary text are in their comparative infancy (de Beaugrande, 1992) research has ensured that the reading process is thought of as not simply a matter of recall of knowledge, but as a 'complex cognitive and affective transaction involving text-based, reader-based, and situational factors.'(Goetz et al. 1993:35)

Text Based factors

Most humanities scholars would agree that their primary task is to determine how meaning can be attributed to texts. (Dixon et al. 1993) Yet the connection between meaning and language is an extremely complex problem. As Snelgrove (1990) concluded, when we read a literary text we understand it not only in terms of the meaning of specific linguistic features, but also by the creation of large-scale patterns and structures based on the interrelation of words and ideas to one another. This pattern making means that the relative meaning invested in a word may depend on its position in a text and the reaction that it may already have evoked in the reader. Dramatic irony, for example, is effective because we know that a character's speech is loaded with a significance they do not recognise. Our recognition of this will depend on the mental patterns and echoes it evokes.

Language may become loaded and suffused with meaning specifically by relations, or use in certain contexts.

The complexity of the patterns generated by human readers is, however, difficult to replicate when computational analysis is used. Text analysis software tends to remove the particular phenomenon under investigation from its immediate context except for the few words immediately surrounding it. A linguist may collect instances of a particular phenomenon and present the results of a concordance sorted alphabetically, irrespective of the order in which the words originally occurred in the text, or the author of them. However for a literary critic the patterns created are vital to the experience of reading the text, and to the way it becomes meaningful. Thus the fragmented presentation of a computer analysis programme cannot begin to approach the kind of understanding of meaning that we gain by reading a word as part of a narrative structure.

The way we read literary text also depends on the genre of the work.

Comprehension depends on what we already assume about the type of text we recognise it to be. Fish, (1980:326) found that he could persuade students that a list of names was a poem because of their assumptions about the form that poems usually take. Hanauer (1998) has found that genre affects the way readers comprehend and recall text, since they are likely to read more slowly and remember different types of information in a poem. Readers also decode terms and anaphora differently depending on what we know to be the genre of a text, for example we will expect 'cabinet' to mean one thing in a newspaper report of a political speech, and another in a carpentry magazine. (Zwaan, 1993:2).

Reader Based factors

The way that we extract meaning from a text also depends on many things which are extrinsic to it. The meaning of language changes depending on the associations an individual reader may make with other features of the text (Miall, 1992), and with other texts and ideas. The creation of such webs of association and causation is central to the historian's craft. As the eminent historian G R Elton put it:

G.M. Young once offered celebrated advice: read in a period until you hear its people speak The truth is that one must read them, study their creations and think about them until one knows what they are going to say next.

(Elton, 1967, p.30)

Interaction between the reader and the text is also affected by factors which are particular to the individual reader. (Halasz, 1992) Iser (1989) argues that narrative texts invite the interaction of the reader by indeterminacy in the narrative. Where there are gaps of uncertainty the reader fills them using their own experience. Readers' experience of a fictional text will even be affected by the relationship which they build between themselves and the narrator. (Dixon and Bertolussi, 1996)

Situational factors

The reader's response to a text is likely to be affected by situational factors, for example their gender, race, education, social class and so on. This is also liable to change over time, so that we may experience a text differently at the age of 56 than at 16. As Potter (1991) argues, these factors have yet to be taken into account in empirical readership studies of literary text. Even more subtly than this however, a reader's appreciation of a text may be affected by how they feel on a particular day, or if a feature of the text is reminiscent of their personal experience. Happy readers notice and recall parts of a text which describe happiness, or evoke the emotion in them, and sad ones the opposite. (Kneepkins and Zwaan, 1994:128) The role of emotional engagement is clearly vital in literary reading. Yet it is one which is very difficult to quantify, describe, and therefore is almost impossible for computer analysis to simulate.

Reading a text is also affected by the frequency of reading and the expertise of the reader, as Elton's observation suggests. (Dixon and Bertolussi 1996 b, Dorfman 1996.) Dixon et al. (1993) found that the same textual feature might be the cause of varied effects in different readers and certain effects were only apparent to some readers. Although core effects of the text could usually be discerned on first reading, other more subtle effects were only reported on second or subsequent readings, or would only be apparent to some of the readers. They also found that the subtlest of literary effects tended to be noticed by readers who they called 'experienced'. They concluded that reading is such a complex procedure that all the effects of the text are unlikely to be apparent at once, and that reading is clearly a skill that needs to be learnt and practised.

Reading the visual

It should therefore be becoming clear why print resources have continued to coexist with digital ones. The key activity of the humanities scholar is to read and interpret texts, and there is little point in using a computational tool to replicate what human agency does best in a much less complex and subtle manner. Reading a printed text is clearly a subtle and complex analysis technique. It is therefore not surprising that scholars have rejected the assumption that digital resources and computational techniques which simply replicate the activity of reading are a pale imitation of an already successful technique. To be of use to the humanities scholar, it seems that digital resources must therefore provide a different dimension that may change the way that we view our raw materials.

In some ways we can discern a similar movement in humanities computing to that which has taken place in computer science. In the 1980s and 1990s artificial intelligence seemed to offer the prospect of thinking machines (Jonscher, 1999, chapter 5). But the technology that has captured the imagination of users has not been a computer systems that seeks to think for them, but one that provides access to material that can provide raw material for human thought processes, that is the internet and World Wide Web . The popularity of the web appears to have dated from the development of graphical browsers that gave us access not only to textual information, but to images. The effect of this and the rise of the graphical user interface has been to reacquaint us with the power of images, not only as ways of organising information, but as way of communicating it. Just as the

images in the museum in Thessaloniki reminded me that there are other ways to interpret and communicate ideas, so we have had to relearn ways to read an image, whether the frame contains a painted or a pixelated icon.

It is in this area that, I would argue, that digital resources can make the greatest contribution to humanities scholarship. Digitisation projects have revolutionised our access to resources such as images of manuscripts, (Unsworth, 2002). The use of three dimensional CAD modelling has been extensively used in archaeology to help reconstruct the way that buildings might have looked (Sheffield University, 2003) However, the projects that are most innovative are not those that use digital resources for reconstruction or improved access, though these are of course enormously valuable, but as tools to think with. If the process of reading and interpreting a text is so complex, then it may be that this is best left to our brains as processing devices for at least the medium term. It is, however, in the realm of the visual that we are seeing some of the most interesting interrelationships of print scholarship and digital resources. We need only look at a small sample of some of the papers presented at this year's Association for Literary and Linguistic Computing- Association for Computers and the Humanities conference (<http://www.uni-tuebingen.de/zdv/zrkinf/pics/aca4.htm>) to see exciting examples of such developments in progress.

Steve Ramsay argues that we might 'remap, reenvision, and re-form' a literary text (Ramsay, 2002). He refers to McGann and Drucker's experiments with deforming the way a written text on a page (McGann, 2001), but has moved beyond this to the use of Graph View Software. He has used this program, which

was originally designed to create graphic representations of numerical data, to help in the analysis of dramatic texts. In a previous interview, he had demonstrated this to me, showing a three dimensional graphical mapping of Shakespeare's *Antony and Cleopatra*. This created wonderful abstract looping patterns which might have been at home in a gallery of modern art. But, like the Macedonian icons, these were not simply objects of beauty. Once interpreted they show the way that characters move through the play, being drawn inexorably towards Rome. This visual representation had the effect not of abolishing the human agency of the literary critic, but providing, literally, a new vision of the play, perhaps opening up new vistas to the critical view. The significance of such movement, and what it reveals about the play, is for the critic herself to decide, but the program has performed a useful form of defamiliarisation, which would be difficult to imagine in a print environment.

We can see similar types of visual representation of textual information in the interactive 3D model of Dante's *Inferno* or TextArc. The effect of both of these is a similar kind of defamiliarisation. A very new view of the information in the text is created, but the effect of it, at least on this reader is to make her wish to return to the text itself in printed form, and to read it with new eyes. The digital resource has therefore not made reading redundant, but helped to suggest new avenues of interpretation. Both of these projects are being developed at the same research centre, IATH where McGann is and Ramsay was based. This is another intriguing connection between the world of digital resources and more traditional forms of scholarship. Computational tools as simple as email have made the process of scholarly collaboration over large physical distances much easier than before. Yet

it is fascinating to note that the physical proximity of scholars such as those at IATH facilitates the interchange of ideas and makes it possible for methodologies to be shared and for projects and scholars to be a creative influence on each other; a process that we can see at work in the visual dynamics of these IATH projects. This is not an isolated phenomenon. The fact that Microsoft's campus in Cambridge shares alternate floors of a building with the University department of computer science shows that such a technologically advanced organisation still values informal creative exchanges as a way to inspire new projects and support existing ones. The Cambridge University Computer Laboratory's online coffee machine was a star turn of the early web. (Stafford-Fraser, 1995) But it was finally switched off in 2001, perhaps proof that Microsoft's decision to privilege meetings over a non-virtual coffee shows their recognition that traditional methods are still vital in innovation and scholarship.

Two other IATH projects were also discussed at the conference. The Salem Witch trials and Boston's Back Bay Fens are two projects which both make use of GIS technology to integrate textual material and numerical data with spatial information. (Pitti et al, 2002) Once again these projects allow the user to visualise the data in different ways. Connections might be made between people or places in historical Boston whose physical proximity is much easier for a user to establish in the visual environment of a GUI interface than by the examination of printed data. Where a particular user's knowledge might be partial, the use of such data lets her literally envision new ones, as a result of interrogating a large special data set. A new textual narrative can emerge from the physical linkages.

The case of the Salem witch trials is also intriguing, since the flash animations actually allow the user to watch as the accusations spread over time and space like a medical, rather than psychological, epidemic. This idea of mapping this spread is, however, not new in terms of historiography. In 1974 Boyer and Nissenbaum had pioneered this approach in a ground breaking book, *Salem Possessed*. This contains printed maps of the area which give us snap shots of the progress of the allegations. We can, therefore, see an immediate relationship between scholarship in print and a digital resources which has grown out of such theories. What the digital resource adds though is the immediacy of being able to watch, run and re-run the sequence in a way that a book, while impressive, cannot allow us to do. Once again, the theories that lie behind the data are, as Pitti et al (2002) make clear, very much the product of the scholarship that the historians who initiated the projects brought to them. Analysis performed on the data will also be done in the minds of other scholars. But the visual impact of both of these databases supports human information processing, and may suggest new directions for human analysis to take.

As the website for The Valley of the Shadow, one of the two original IATH projects, puts it, GIS may be used literally 'to put [historical] people back into their houses and businesses'. (Ayers, et al. 2001) Valley of the Shadow is itself doing far more than simply this. The project itself seems to be predicated on an understanding of the visual. Even the basic navigation of the site is organised around the metaphor of a physical archive, where a user navigates the different materials by visiting, or clicking on, the separate rooms of a plan of the space. By linking a contemporary map with the data about a particular area GIS allows users

to interpret the data in a new way, giving a concrete meaning to statistical data or names of people and places in the registers which are also reproduced. (Thomas, 2000) Just as with a literary text, this might cause a historian to return to the numerical or textual data for further analysis, and some of this might use computational tools to aid the process.

But its greatest value is in prompting a fresh approach for consideration. The historian's brain is still the tool that determines the significance of the findings. This distinction is important, since it distinguishes Valley of the Shadow from some earlier computational projects in similar areas of American history. For example, in 1974 Fogle and Engerman wrote *Time on the Cross*, in which they sought to explain American slavery with the use of vast amounts quantitative data on plantation slavery, which was then analysed computationally. This, they claimed would produce a definitive record of the objective reality of slavery in America. Critics of the book have argued persuasively that the data was handled much too uncritically. Their findings, it has been claimed, were biased, because they only used statistical written data, which usually emerged from large plantations, and ignored the situation of small slave holders who either did not or could not document details of their holdings, either because the holdings were small or because the slave holder was illiterate. (Ransom and Sutch, 1977, Wright, 1978) Other historians have insisted that anecdotal sources and printed texts must be used to complement the findings, to do sufficient justice to the complexity of the area. It could be argued that the problems that they encountered were caused by an over reliance on computational analysis of numerical data, and by the implication that this could somehow deliver a definitive explanation slavery in a

way that would finally put an end to controversies caused by subjective human analysis. A project such as Valley of the Shadow is a significant progression onwards, not only in computational techniques but also scholarly method. It does not rely on one style of data, since it links numerical records to textual and spatial data. These resources are then offered as tools to aid interpretation, which takes place in the historian's brain, rather than in any way seeking to supersede this.

The products of the projects are also varied, ranging from students projects, which take the materials and use them to create smaller digital projects for assessment, to more traditional articles and conference presentations. Most intriguing, perhaps is the form of hybrid scholarship that Ed Ayers, the project's founder, and William Thomas have produced. Ayers had long felt that despite the innovative research that could be performed with a digital resource, there had been very little effect on the nature of resulting publication. The written article, even if produced in an electronic journal was still essentially untouched by the digital medium, having the same structure as an article in a traditional printed journal. Ayers and Thomas (2001) therefore wrote an article which takes advantage of the electronic medium, by incorporation some of the GIS data, and the hyper textual navigation system which gives a reader multiple points of entry and of linkage with other parts of the resource. Readers are given a choice of navigation, a visual interface with flash animations or a more traditional text based interface. The article might even be printed out, but it is difficult to see how it could be fully appreciated without the uses of its digital interactive elements. This will appear in *American Historical Review*, a traditional academic journal, announcing its right to be considered part of the mainstream of historical research. As such is represents a dialogue between

the more traditional world of the academic journal and the possibilities presented by digital resources, at once maintaining the continuity of scholarly traditions in history, but also seeking to push the boundaries of what is considered to be a scholarly publication. The analysis presented in the paper emerges from human reading and processing of data, but would not have been possible without the use of the digital resource.

It is not only at IATH, however, that work in this area is taking place, despite their leading position in the field. Thomas Corns, (2002), from the University of Bangor in Wales has also described how one aspect of document visualisation can aid human analysis. Being able to digitise a rare manuscript has significantly aided his team in trying to determine whether it was written by Milton. The simple task of being able to cut, paste and manipulate letter shapes in the digitised text has helped in their examination of scribal hands. The judgement is that of the scholars, but based on the ability to see a text in a new way, only afforded by digitised resources. This is not a new technique, and its success is largely dependent on the questions that are being asked of the data by the human investigators. Donaldson, (1997) discusses ways in which complex analysis of digital images of seventeenth century type was used to try to decide whether Shakespeare had used the word wife or wise in a couplet from the *Tempest*, usually rendered as 'So rare a wondered father and a wise/Makes this place paradise' (Act IV, Scene, I, ll. 122-3). The digital research proved inconclusive but might have been unnecessary, since a Shakespeare scholar might be expected to deduce that the rhyme of wise and paradise is much more likely in the context of the end of a character's speech, than the word wife, which while tempting for a

feminist analysis would not follow the expected pattern of sound. All of which indicates that the use of digital resources can only be truly meaningful when combined with old fashioned critical judgement.

Another project being presented at ALLC-ACH, which is very much concerned with facilitating critical judgement though the realm of the visual, is the Versioning Machine (Smith et al., 2002). This package which supports the organisation and analysis of text with multiple variants is once again a way of helping the user to envision a text in a different way, or even in multiple different ways. The ability to display multiple variants concurrently, to colour code comments that are read or unread, selectively to show or hide mark-up pertaining to certain witnesses, gives scholars a different way of perceiving the text, both in terms of sight and of facilitating the critical process. It is also far less restrictive than a printed book in the case where the text of a writer might have multiple variants, none of which the critic can say with certainty is the final version. The case of Emily Dickinson is a notable one, that is presented by the MITH team, but it may be that if freed by digital resources like the Versioning Machine of the necessity of having to decide on a copy text for an edition, the text of many other writers might be seen as much more mutable, and less fixed in a final form of textuality.

Print editions, for example of the seventeenth century poet Richard Crashaw have forced editors to make difficult decisions about whether the author himself made revisions to many of the poems. When, as in this case, evidence is contradictory or inconclusive, it is surely better to be able to use digital technology such as the

Versioning Machine to give users a different way of seeing, and enable them to view the variants without editorial intervention. The use of the Versioning Machine will not stop the arguments about which version might be preferred, based as they are on literary judgement and the interpretation of historical events, but at least we as readers are not presented with the spurious certainty that a print edition forces us into. Once again, therefore, such use of computer technology is not intended to provide a substitute for critical analysis, and the vast processing power of the human brain, rather it gives us a way of reviewing the evidence of authorial revisions. It makes concrete and real again the metaphors that those terms have lost in the world of print scholarship.

Conclusion

Even when we look at a small sample of what is taking place in the field, it is clear that some of the most exciting new developments in the humanities computing area seem to be looking towards the visual as a way of helping us to reinterpret the textual. It appears that we are moving beyond not printed books, and print based scholarship, but the naïve belief that they can easily be replaced by digital resources.

As the example of my visit to Portsmouth demonstrated, it is simplistic to believe that we can, or should rush to convince our more traditional colleagues of the inherent value of digital resources, without taking into account the culture of long established print scholarship. It is only through negotiations with scholars, and in

forging links between the digital and the textual traditional that the most interesting research work is likely to emerge.

The materials that humanities scholars use in their work are complex, with shifting shades of meaning that are not easily interpreted. We are only beginning to understand the subtle and complicated processes of interpretation that these require. However, when we consider that the process of reading a text, which may seem so simple, is in fact so difficult an operation that computer analysis cannot hope to replicate it at present, we can begin to understand why for many scholars the reading of such material in print will continue to form the core activity of their research.

Digital resources can, however, make an important contribution to this activity. Far from attempting to replace the scholar's mind as the processing device, computer delivery of resources can help to support the process. The complexity of visual devices as a way of enshrining memory and communicating knowledge is something that the ancient world understood very well, as I learnt when I began to read the icons in Thessaloniki. While much of this knowledge has been lost in the textual obsession of print culture, the graphical interface of the computer screen has helped us reconnect to the world of the visual and recognise that we can relearn a long neglected vocabulary of interpretation. Digital resources can provide us with a new way to see, and thus to perceive the complexities in the process of interpreting humanities materials. A new way of looking at a text can lead to a way of reading it that is unconstrained by the bindings of the printed medium, even if it leads us back to the pages of a printed book.

CLAIRE WARWICK, University College London. December 2002.

Bibliography

Ayers, E. L. (1999) The pasts and futures of digital history.

<http://jefferson.village.virginia.edu/vcdh/PastsFutures.html>

Ayers, E. L. Rubin, A. S., Thomas, W. G. (2001) The Valley's Electronic Cultural Atlas Initiative geographic information systems project.

<http://jefferson.village.virginia.edu/vshadow2/ecai/present.html>

Ayers, E.L. and Thomas, W. G. (2002) Two American communities on the eve of the civil war: an experiment in form and analysis.

http://jefferson.village.virginia.edu/vcdh/xml_docs/projects.html

de Beaugrande, R. (1992). Readers responding to literature: Coming to grips with reality. In Narduccio, Elaine F. (Ed.) *Reader Response to Literature: The Empirical Dimension*. (pp. 193-210). New York and London: Mouton de Gruyter.

Boyer, P. and Nissenbaum, S. (1974). *Salem Possessed : the Social Origins of Witchcraft*: Cambridge, Mass, London: Harvard University Press.

Burrows, J.F. (1987). *Computation into Criticism: a Study of Jane Austen's Novels and an Experiment in Method*. Oxford, Clarendon Press

Computers and the Humanities, 27, (1993). Special issue on literature and computing, includes: Thomas, J-J. (1993). Texts on-line. 93-104. Bruce, D. Towards the implementation of text and discourse theory in computer assisted textual analysis, 357-364. Fortier, P.A. Babies, bathwater and the study of

literature. 375-385. Lessard, G. and Benard, J. Computerising Celine. 387-394.
Olsen, M. Signs, symbols and discourses: a new direction for computer aided
literature studies. 309-314. Olsen, M. Critical theory and textual computing:
comments and suggestions. 395-400.

Corns, T. N. (1990). *Milton's Language*. Oxford: Blackwell.

Corns, T.N. (1991) Computers in the humanities: methods and applications in the
study of English Literature. *Literary and Linguistic Computing* 6, (2) 127-131

Corns, T. N. et al. (2002). Imaging and amanuensis: understanding the manuscript
of *De Doctrina Christiana*, attributed to John Milton. Proceedings of ALLC/ACH
2002 New Directions in Humanities Computing. The 14th International
Conference. University of Teubingen, 24-28 July 2002. pp. 28-30.

Crowder, R.G. and Wagner, R.K. (1992). *The Psychology of Reading: An
Introduction*. Oxford: Oxford University Press. (2nd edition)

Dixon, P. et al. (1993). Literary processing and interpretation: towards empirical
foundations. *Poetics* 22, 5-33.

Dixon, P. and Bertolussi, M. (1996). Literary communication: effects of reader-
narrator co-operation., *Poetics* 23, 405-430.

Dixon, P. and Bertolussi, M. (1996b). The effects of formal training on literary reception. *Poetics* 23, 471-487

Donaldson, P. S. (1997). Digital archive as expanded text: Shakespeare and electronic textuality. In Sutherland, K. (Ed). *Electronic Text: Investigations in Method and Theory*. (pp. 173-198) Oxford: Clarendon Press.

Dorfman, M.H. (1996). Evaluating the interpretive community: evidence from expert and novice readers. *Poetics* 23, 453-470.

Duff, W. M., and Johnson, C. A. (2002). Accidentally found on purpose: information-seeking behavior of historians in archives. *Library Quarterly*, 72, 472-496

Elton, G.R. (1967). *The Practice of History*. London: Fontana.

Finneran, R.J (1996). *The Literary Text in the Digital Age*. Ann Arbor: University of Michigan Press.

Fish, S. (1980). *Is there a Text in this Class? The Authority of Interpretive Communities*. Cambridge Mass: Harvard University Press.

Fogel, R. W. and Engerman, S. L. (1974) *Time on the cross*. London : Wildwood House.

Goetz, et al. (1993). Imagery and emotional response. *Poetics* 22, 35-49.

Halasz, L. (1992). Self-relevant reading in literary understanding. In Narduccio, Elaine F. (ed.) *Reader Response to Literature: The Empirical Dimension*. (pp. 229-246). New York and London: Mouton de Gruyter.

Hanauer, D. (1998). The genre-specific hypothesis of reading: Reading poetry and encyclopaedic items. *Poetics* 26, 63-80.

Iser, W. (1989). *Prospecting: From Reader Response to Literary anthropology*. Baltimore and London: Johns Hopkins Press.

Jonscher, C. (1999). *The Evolution of Wired Life: From the Alphabet to the Soul-Catcher Chip—How Information Technologies Change Our World*. New York: Wiley.

Kneepkins, E.W.E.M. and Zwaan, R.A. (1994). Emotions and literary text comprehension. *Poetics* 23, 125-138

McGann, J. (2001). *Radiant Textuality: Literature after the World Wide Web*. New York and Basingstoke: Palgrave.

Miall, D. (1992). Response to poetry: studies of language and structure. In Narduccio, Elaine F. (Ed.) *Reader Response to Literature: The Empirical Dimension*. (pp. 153-172). Berlin and New York: Mouton de Gruyter.

Norman, D.A.(1999). *The Invisible Computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution.* Boston: MIT Press.

Nunberg, G. (ed.) (1996). *The Future of the Book.* Berkeley: University of California Press. p.10

Pitti, D. Jessee, C. and Ramsay, S. (2002). Multiple architectures and multiple media: the Salem witch trials and Boston's Back Bay Fens. Proceedings of ALLC/ACH 2002 New Directions in Humanities Computing. The 14th International Conference. University of Teubingen, 24-28 July 2002. pp. 87-91.

Potter, R.G.(1991). Pragmatic research on reader responses to literature with an emphasis on gender and reader responses. *Revue Belge de la Philosophie et d'Histoire.* 69, 3, 599-617.

Ramsay, S. (2002). Towards an algorithmic criticism. Proceedings of ALLC/ACH 2002 New Directions in Humanities Computing. The 14th International Conference. University of Teubingen, 24-28 July 2002. p. 99.

Ransom, R. L., Sutch, R. (1977) *One kind of freedom : the economic consequences of emancipation.* Cambridge : Cambridge University Press.

Richards, I.A. (1929). *Practical Criticism: A Study of Literary Judgements*.
London: Keegan Paul.

Sheffield University (2003) *The Cistercians in Yorkshire*.
<http://cistercians.shef.ac.uk/>

Smith, M. N. (2002). MITH's lean, mean versioning machine. *Proceedings of ALLC/ACH 2002 New Directions in Humanities Computing. The 14th International Conference. University of Teubingen, 24-28 July 2002. pp. 122-127.*

Snelgrove, T. (1990). A method for the analysis of the structure of narrative texts. *LLC 5, 221-225.*

Stafford-Fraser, Q. (1995) *The Trojan Room coffee pot: a (non-technical) biography*. Cambridge University Computer Laboratory.
<http://www.cl.cam.ac.uk/coffee/qsf/coffee.html>.

Stone, S. (1982). Humanities scholars-information needs and uses. *Journal of Documentation. 38 (4), 292-313*

Thomas, W. G. (2000) *Preliminary conclusions from the Valley of the Shadow Project*. Paper delivered at: ECAI/Pacific Neighborhood Consortium Conference, January 11, 2000, University of California at Berkeley.
<http://jefferson.village.virginia.edu/vcdh/ECAI.paper.html>

Unsworth, J. (2002) *Using digital primary resources to produce scholarship in print*. Paper presented at [Literary Studies in Cyberspace: Texts, Contexts, and](#)

[Criticism](#). Modern Language Association Annual Convention, Sunday, December 29, 2002. New York, USA. <http://jefferson.village.virginia.edu/~jmu2m/cyber-mla.2002/>

Van Peer, W. (1989). Quantitative studies of literature, a critique and an outlook. CHUM, 23, 301-307.

Wright, G. (1978) The political economy of the cotton South : households, markets, and wealth in the nineteenth century. New York : Norton.

Zwaan, R.A. (1993). Aspects of literary Comprehension: a Cognitive Approach. John Benjamins: Amsterdam.

ⁱ It is with gratitude that I would like to dedicate this chapter to the memory of Dr J. Wilbur Sanders 1936-2002. A man of acute perception and a great teacher, he helped me to see how subtle and complex a process reading might be.