Leicester workshop 2007

[title slide]

1. Digital Revolution?

Work in digital humanities exists at an intersection between disciplines which traditionally have known little of each other – perhaps even been suspicious or disparaging about each other. The uncertainly can create polarisation where may people exhibit a tendency to be either overly positive or negative about digital media and the potential of digital resources to revolutionise the way we do our research and teaching While some talk about 'digital revolution' and speak of digital resources as if they are to humanities scholars huddling in their cardigans in dusty archives as they pore reams of crumbling paper what the invention of the washing machine or vacuum cleaner promised to be for domestic life - others are resistant and exhibit a kind of snobbery about these new media. Neither of these positions is particularly helpful to us. There are many gains to be made but there are also pitfalls and costs involved and we need to make choices critically. This needs a balanced view which can only come from knowledge about the field.

[slide]

2. What it will do

- The searchability of electronic text opens it up to new ways of doing research
- Marking up texts and adding metadata allows these objects to contain different orders of information and inscribes relationships into the objects.
- This applies to both the scholarly editions of nineteenth-century objects (exhibitions, book, websites) and the secondary literature that we produce about them (journal articles, descriptions, pedagogic material)
- But most importantly, digital technologies allow us for the first time to republish certain orders of material – first editions, rare bindings, pictures,

manuscripts and crucially unwieldy bulky paper objects such as periodicals and journals. Mention ncse *NS* – most complete in the world.

- This could make hitherto rare, fragile objects as visible as well-known novels and poems, changing the cultural landscape of the nineteenth century.

This process is underway: our discipline is about to change. For instance there are at least three large scale projects underway which will publish digital editions of hundreds of nineteenth-century periodicals. Even if you aren't interested in the periodical press as such, the availability of so much information will inevitably recontextualize certain events and figures.

[slide]

3. Why should we care?

- Distinction. In order to preserve the identity and authority of research that we do in the humanities we have to develop sophisticated ways of using digital resources which set it apart from simple Googling. We need to find ways of integrating the research and teaching methods and standards which we have evolved over hundreds of years of solid scholarship with research and teaching in the digital domain.
- We also need to consider our involvement with the process of creating resources so that the fruits of our scholarship – and the publicly owned resources on which they are based are not simply assimilated and then sold back to us by large publishing companies and repositories at a high premium.
- We also have to think carefully about how the proliferation of digital resources can reconfigure the cannons of our discipline and so the preoccupations of academic study. If we are not involved in selecting and designing the content of these resources we face using resources which might mask or obscure what scholarship deems to be important aspects of our cultural heritage. There are also questions to be engaged with around appropriate amounts of intervention in digital texts and whether we should be pursuing an archive or an edition model

2

in republishing this material today. (We'll come on to talk about this some more in the last part of the workshop.)

- We also need to think about this in terms of shaping future academic careers- not only with how it shapes our day to day work.

Funding: academics are now perhaps more likely to get a large sum of money for creating a digital resource than writing a book because creating a resource that will be open access and of public interest is a more justifiable use of public money but this means that there is pressure on academic projects to address audiences outside the academy.

In the UK the funding bodies shape the research produced by academics. The AHRC is increasingly results-focused, operating more as an academic press than a funding body. The results of this are:

- Small scale research projects, usually completed in three years with small amounts of (non-renewable) funding
- These projects tend to be stand-alone, although JISC are working on portals that will allow interoperability.
  http://www.jisc.ac.uk/index.cfm?name=programme_portals
- Diffusion of expertise after projects come to an end, and lack of forums for discussion and collaboration. Attempts to prevent / consolidate this through AHRC ICT methods network.
- the projects have an uncertain future: some universities charge for hosting resources; but AHDS (joint JISC and AHRC) offers a home for a copy.

There are other funding streams:

- joint AHRC, JISC, EPSRC research fund (2m pounds, including 4 postgrad studentships). This brings humanities within the scope of e-science research. 48 months in duration, and for between £20,000 and £400,000 (fEC) – which is pretty rubbish when you remember that a postdoc salary will eat at least 20,000

a year- plus overheads. And these (like other AHRC grants) must be collaborative and you must have an institution.

- Leverhulme has research project grants (up to 1.75 million), as well as its more general funding for research.

So the funds are there, especially if you are willing to play the game of the funding bodies strategic objectives and this also means that many of the postdocs available for young scholars are working as research assistants on digital projects.

4. Postdoc work in the digital domain.

Us: Jim and I are both postdocs on the NCSE project. Me: background (psychology, philosophy, history, and now computing and print culture).

When I started my first post doc – I had quite an insular view of the project (large book-history orientated bibliographical database of 19[th]C publication) - not aware that I was also traversing a border into a sometimes strange foreign lands of digital humanities or that this was a burgeoning field of scholarship in its own right. I've been working on digital projects in one form or another ever since.

Some of the benefits of being involved in digital projects:

Intellectual issues- Its not all nerdy computing and grunt technical work - as I've hinted there are interesting intellectual issues around doing any humanities work digitally. As well as dealing with all the familiar areas of enquiry associated with their discipline scholars involved in doing digital projects have to engage with issues around:

- The complex relationships between archival materials and their digital surrogates – this often involves debate about representation which can go from practical and techy to rather lofty and philosophical. We'll look at some examples from ncse in a minute. Digital humanities places representation of archival materials on a par with critical analysis or enquiry- and we hope to stimulate discussion on this.

- As I've mentioned achieving a synthesis between the interpretive strategies employed by scholars in the humanities and the digital tools created to facilitate exploration and analysis of original artefacts is another concern.
- Relatedly we have to understand the tools being used and the implications of decisions we make in the use and creation of those tools and their impact on our analytical processes.

New ways of working- involvement in digital projects changes the kind of collaborative partnerships we have to be involved with to produce scholarly outputs- we often find ourselves working with commercial or non-academic public sector partners who have very different imperatives than us and with computing professionals as well as colleagues in other disciplines. Teams can be larger and more scattered than is usual in some more traditional forms of academic work and this means that more effort needs to be put into coordinating collaborative work and making sure that all the parties understand each other. However, it also means that you get a very varied experience and build relationships and knowledge in areas that might be outside your discipline. This necessary interdisciplinary can be both a positive and a negative for postdocs in terms of their future employability depending on whether employers feel it broadens expertise or spreads it too thinly across too wide an area.

There are also issues around publication. Collaborative work often means collaborative research outcomes. Although this has long been an acceptable model in the sciences, in the arts co-authored books and articles are still often seen as having less weight than the monograph or single author article.

This and other issues also affect digital publications specifically. Academic esteem, and thus employment chances are partly defined in terms of publication in well established peer reviewed journals –and researchers can spend many years putting together a digital publication which has less prestigious status for the RAE than a monograph or journal articles. This is changing slowly and researchers in humanities computing are looking at ways to give electronic publications equivalent status through peer reviewing (*19* e.g.) but this change will take time and in the meantime young scholars involved in digital

publication also need to produce print publications. There are also issues around stability-the project's funding may only cover hosting for a limited period – a digital resource is not an object you can hold in your hand or own and point to on a library shelf. The longevity of your research outcomes are subject to continuing funding for hosting and upgrading as necessary and continuing use and relevance. Collaborative working also means that the credit for these outputs and any ownership of them is distributed as as a post doc you might not own the IP rights to your work.

Institutional benefits (or not!) – (JM) I'm employed as 0.5 fte on an AHRC resource enhancement grant for 3 years. This is great as it means that I have access to institutional support (office space, photocopying and printing, email and internet access, some funding for personal research, access to colleagues) – something I do not always get as part of my part-time teaching. However, it is all pro rata, making it difficult to attend conferences etc, or find time to do any of my own research.

[slide]

5. ncse

intro:

- Digital edition of six periodical titles

    o Monthly Repository 1806–1838 (plus its supplement, the Unitarian Chronicle and including its multiple editions)
    o Northern Star 1837–1852 (including the Chartist portraits published as supplements and including its multiple editions)
    o Leader (1850-1860)
    o Tomahawk (1867-1870)
    o English Woman's Journal (1858-1864)
    o Publishers' Circular (1880-1890)

[slide]

why is it special:

- It is a collaboration between BL, CCH, Olive software, but will produce a freely accessible online resource.
- We are working with Olive Software and CCH to develop ways of identifying and capturing information beyond simply relying on OCR to produce an index.
- Developing metadata structures to allow comparisons across the titles. These include indices of people, places, events, institutions, and publications, and data mining to provide subject indexes which we will link to thematic concepts.
- It is not just a delivery system that serves an archive of content, but rather an edition which foregrounds the contextual and referential nature of C19th print culture.

[slide of BL Olive pilot] (Archives and editions)

[slide]

Some editing dilemmas:

(TH- two main issues- reproducing levels of meaning in image and reproduction of images in colour)

<u>Tomahawk and the muck-rake</u>

Each number of Tomahawk featured a cartoon, on a separate sheet of paper, which accompanied a "Sketch" in the paginated letterpress. *Tomahawk* was a satirical rival to *Punch*, selling around 50,000 copies per week. The title cast itself as standing for truth, and the eponymous native American would "scalp" those it felt were misrepresenting the affairs of the day. The image has *Tomahawk* overlooking the workings of the dissolute "Penny Press" (it says who he is in his hat) as he rakes the muck with his "Sensational Claptrap." Unseen behind the "Penny Press" is an Angelic female figure who is carries the crown of "Truth." A motto – just below the subtitle in the original, I'll explain why it isn't in the slide shortly – relates the image to *Pilgrim's Progress*.

These words, operating at the level of iconography, and participating in the organizational logic of the image, are also unlikely to be captured by OCR processes, and so would not be "searchable" by users. The only aspects of the image that would be accessible by text searching is the caption but, without the accompanying image, its meaning is quite oblique. For instance, although "People's Guide" signals its relevance towards working class politics, there is nothing in the title or the motto to signal that the image is about the press. The cartoon also refers to *Tomahawk's* role within print culture. Although it was only tuppence weekly, *Tomahawk* was not intended for the "people:" its gentlemanly satire requires knowledge of metropolitan high society, and it had no pretensions to speak to (or indeed for) the working class. However, it did reserve the right to judge its politics and, as the picture suggests, an entire genre of periodicals. This picture, in other words, is not just a comment on the "Penny Press," but also locates *Tomahawk* in a different sector of the print trade. Any metadata seeking to identify the image would have to also recognize this self-reflexive level of meaning.

Another issue raised by this example is that of the relationship between the hard copy and the digital surrogate we are creating.

[slide]

Whereas as the other images in this paper are taken from the tiff images generated from microfilm prepared for **ncse**, this image is scanned from a photocopy taken from the hard copy. The reason for this is that, in order to get good definition of the black letters on the white paper, the tonal contrast for the microfilm needs to be set quite high. This means that the various grey effects created by close lines in engravings are not captured very well, usually coming across as uniform dark areas.

Although in this case the operators have recognized that images would benefit from different settings and have taken another image, often they do not, showing a clear bias towards the written text. However, as the microfilm technicians cannot see the image that they capture, they can only estimate the amount of contrast necessary and here you can see that they have over-compensated in the lighter image. This problem can be

overcome by inserting digital images that are taken directly from the hard copy. However, as filming each page is time-consuming, libraries like to use microfilm – which has already undergone this process – rather than film once again from hard copy. This makes sense if all you want to do with your digital edition is make use of well-defined text, but such a methodology limits the extent to which one can work with either the original colours on a page – whether in coloured letterpress, coloured images, or on coloured paper – or even the subtle variations that exist even in monochrome source material.

These problems with reproduction reveal the mediating technologies that create the image – in this case it is the older technology of microfilm – and we, as editors, have to compensate for them to return to some sort of original, however we might conceive it. We now have the option to remedy this by including high resolution colour images of a selection of the worst affected of the cartoons in the edition but we have to decide where to include them. Should we replace the cartoons poorly reproduced from microfilm with colour images, interleaving these with bitonal images of the other pages or house them in a separate area of the site?

[end slide]

OPEN FLOOR