Editions and archives: textual editing and the Nineteenth-Century Serials Edition

[slide 1]

1.0 Editing Periodicals in the Digital Domain.

- Freed from the material constraints of paper, digital publication allows us to republish whole runs – and indeed collections of runs – for the first time.
- There has been plenty of discussion about how textual editing in the digital domain varies from its paper-based antecedents. Digital versions of books have been complemented by renderings of gallery spaces, museum exhibitions, and historical buildings.
- However, the recent trend to digitize runs of periodicals employs an unreflective editorial methodology that pays little attention to the formal demands of the genre. Today we want to identify these demands, and consider the ways in which digital tools allow us to translate them into the digital domain.

2.0 The periodical as genre.

The periodical makes specific generic demands however it is usually conceived of as a book that happens to have been published serially. This model does not adequately cover the diversity of periodical content, nor the contextual nature of its contents.

Periodicals differ from books in the following ways:

- People: multi-authored, publishers, editors, printers, illustrators. Without the figure of the author to mask the collaborative nature of publication, all those involved become candidates for a focus.

[slide 2]

- Uncertain beginnings and endings – *Monthly Repository* family tree.
- Uncertain seriality – different rhythms, presence of multiple editions, supplements, prefaces etc. The run is not a straightforward linear sequence like a series of chapters.
- Size of runs: 98,565pp is too many to read, and certainly too many to publish in paper.
- The "unit" of analysis (number? Volume? Year? Article? These are all different objects, within different contexts and audiences. What are their relationships? Volumes are not commensurate with calendar years for instance.)
- They are linked to specific temporal moments: their temporal rhythm implies a more delimited address: i.e. more titles are discussing the same moment.

[slide 3]

There are few editors today that would argue that meaning is entirely divorced from the historical context in which it is composed. When dealing with periodicals, in which the temporal moment of publishing is not only referred to within individual numbers but is also announced on the front cover, the relationship to their immediate historical context is doubly foregrounded. When we remember that these numbers are gathered in runs, sometimes spanning many years, periodical titles can reflect broader cultural shifts in their changes of address, form, and content over the run.

It is important that digital editions of periodicals open up content to users. However, the nature of this content is contextual and therefore we have to address relationships:

- Structural/hierarchical relationships:

  Between different levels of content – between individual articles and departments; between different editions of the same number; between, numbers and supplements and numbers and volumes; volumes and series and between editorships.

- Generic/thematic relationships:

  Between equivalent genres of content across the run of a periodical and between periodicals; subject classifications of content; different configurations of content reflect imagined readers etc.

- Form, typographies:

  Changes in layout within the run of a periodical; shared forms between periodicals; changes in technology alter periodical form.

- Content:

  'Borrowed' content; reporting/commenting on the same news events and issues; circulation of articles and images as part of the market for print.

- Prosopographical relationships:

  Personnel involved in production: possible relationships between printers, publishers, editors, contributors, illustrators, engravers and paper merchants. These personnel relationships map onto broader social groupings.

Summary:

The big question is who you edit these for. Different users have different interests in these relationships.

However, this should be balanced against the question of preservation. The ephemerality of periodicals – both in terms of materiality and discourse – means that digitization is often allied to preservation for journals in a way that it is not for other types of print. The associated difficulties in preserving hard copy (they are often printed on poor quality paper and take up lots of space) provides more pressure to record and archive not just what they say, but also what they look like. Unlike the previous move to shift archives into microfilm, digital preservation is a much more versatile medium that offers us the chance to minimize the loss of information that corresponds to radical transformations of material form.

[slide 4]

2.0   Existing models

- They privilege content, conceiving it as located in stand-alone articles that can be separated from their contexts.
- e.g. British Library *Penny Illustrated Paper*. Not only do users have to specify a keyword (they cannot search by date in order to browse), thereby implying all users want to find articles about something, but when hits are returned they takes users straight to the segment. Users have to choose to select full page view in order to see what else is on the page. In this example the whole page view reveals that a part of the segment is missing, and that this is an advertisement accompanying one for lemonade, a prize competition, the end of a story, and a news report about a retired army officer on trial for embezzlement. Also, as users can only zoom into segments, they cannot see the page numbers, which do not correspond to those given in the display. However this collection is designed for general users and has a simple, easy to use interface.

[slide 5]

This slide shows the contextual relevance of items: this is by O'Connor, the proprietor of the paper, and is in the most important textual position.

- There is historical precedent for models based on content-retrieval. Periodical indices such as *Poole's* have always operated at the level of the article – this model doesn't represent context (either historical context, or immediate context on the page).
- But there other models: for instance the press directories (which are intended to let advertisers know about different periodical titles) are at level of title!

Summary: the presence of these other models of indices reminds us that there are other ways of conceiving of periodicals than as simply repositories of articles. The Press Directories testify to the importance of the title's identity, something that has been taken up by recent bibliographic projects the Wellesley Index and the Waterloo Directory. We argue that the recent attention to print forms that have been marginalized in the past is a salutary reminder that the material culture of print is as worthy of editorial attention as

what is written within them, and this is especially the case in digital republication, which involves such radical violence to material form. Form – whether conceived in terms of paper, typography, spatial layout, size, or genre – not only ensures the passage of texts through time, but also embodies the interaction between a text and its immediate context, inscribing both its processes of production and its intended future life onto the page.

3.0   ncse is a digital <u>edition</u>.

[slide 6]

- In ncse we aimed to produce an edition of six titles, which would represent the interconnected nature of C19th print culture. As an AHRC-funded project, with limited funds and lifetime, we cannot devote resources to preserving large tracts of content. This is done by private companies such as Proquest / Thomson Gale or national institutions such as the BL.
- There are reasons for including each title (MR early poetry / dissenting culture; NS Chartism; EWJ feminism; PC print trade; Tomahawk political satire / cartoons; L intellectual radicalism), and we want to re-publish them for users today. This means we want to provide access to the titles themselves as historically significant objects, rather than simply as repositories of articles.
- Although there are broad similarities between the six titles – there is a bias towards reform etc… But their selection is not intended to make an argument about their relationship, but rather exemplifies the following:
    - Diversity of periodical forms
    - Diversity of material (in terms of content)
    - Challenge distinction between newspaper and periodical.

[slide 7]

Summary of this point: although they appear diverse [and explain slide], what they share are generic features, and a common origin – they are products of the C19th print trade. This is why it is important to make digital resources do more than provide OCR transcripts as indices to preservation tiffs. We need to apply digital tools to the objects themselves, rather than just their letterpress, to bring out their constitutive relationships.

[slide 8]

For instance, paper size links to the economics of production and ideas of readership

[describe slide: digital images obscure the size change – just before it becomes the *Star of Freedom* (p.15 no.753 17 April 1852), the editor, G. Julian Harney, announces his return with a price cut (from 5d to 4½d). He reveals it cannot be any cheaper without having to cut the size in half, or filling it with police reports to maintain the necessary circulation of 30,000 a week. From 14 August 1852 it undergoes an even more radical change, 3 cols and smaller, but twice as long (16pp) – the editor suggests it will encourage people to buy it in ½ yearly volumes. In other words it is no longer a

newspaper, and its format – even to the extent of publishing a town and country edition – resembles the *Leader*.]

They are artefacts that testify to readership and the processes of production, both of which are notoriously difficult to research as there is little surviving evidence for either. Whereas we have commentary for the *Northern Star*, we have little for the *Leader*, so the signification of its format has to be implied from both its form, and its relationships with its competitors.

4.0   Editorial challenges in ncse:

The periodicals in **ncse** are deliberately diverse. However, they still constitute a genre (they are published serially, are composed of discrete units etc…). Our content analysis tools (advanced metadata, text mining, free text searching, concept maps) ensure that content is presented relationally, but what about the formal distinctions within the genre?

[slide 9]

Some journals approach book-like forms:

- EWJ – it really is a series of articles as they are all numbered sequentially throughout the volume – it asks readers to encounter a sequence of articles. It is like a multi-authored book. Part of its appropriation of seriousness.
- But even this title inscribes heterogeneity – for instance "Open Council" and "Passing Events" recur monthly, are dialogic, and are related to the moment.

But the Northern Star approaches the newspaper:

- Newspapers arrange information spatially (synchronically), and layout is therefore meaningful. Its disparate articles all address the same temporal moment, but offer differing perspectives upon it.
- Multiple editions **[slide 10]**. They make things larger and complicated. For instance in the first edition for 20 August 1842, the *Northern Star* went to press with an embarrassing white space on its front page: priority is timeliness but the space reveals the size, retention of the caption reveals what it was, etc etc…

OCR indices are good at text, but miss any non-verbal content. This means we need editorial intervention to make this content available. For instance:

[slide 11]

- Words can be pictures such as in the *Northern Star* and *The Times*.

[slide 12]

- *Leader*: borders, the image of the coffin, "the body" in gothic font.

[slide 13]

- Words can be in pictures *Tomahawk*.

[slide 14]

- seem self-evidently important – like the cartoons in *Tomahawk* – but there is a relationship between the text and the image. Should we translate the reference into a hyperlink?

Summary: the use of OCR to produce transcripts and the display of tiff images is not enough. Although in combination the images compensate for the over-privileging of the words in OCR-generated transcripts, the relationships between these also need to be foregrounded. In a contextual environment such as the periodical page, it is important not to over-determine the division between form and content.

5.0   Representing Form

Olive are good at retaining the connection between form and content by linking OCR transcripts to the segments on the tiff images. As they produce xml repositories, there is also a large potential for further manipulation of these textual units. However:

[slide 15]

- Their applications recognize limited models of print publication. APA does newspapers, Enterprise does books (and conceives of journals as chapters of books).

We've been working with them to merge their applications, but this only approximates the periodical genre, which is not simply a mix of newspaper and book. Providing a top-down abstraction of the genre is very complicated. Working from our six periodicals produced the following data map:

[slide 16]

This result was methodologically complex, and implementing it was beyond the scope of our resources. Instead Olive worked upon sample pages to identify a hierarchy of textual elements based upon standard rules of physical presentation.

[slide 17]

This screen shot shows the segmentation working to four levels. Within the number (itself within a volume and a run), we have "Portfolio," then *The Apprenticeship of Life* (a serial novel), then "First Episode," then the chapter itself "The Young Sceptic." This

hierarchy is captured in the ToC on the left, and the segment is highlighted on the page on the right.

These results are good, but they took a great deal of work from research teams both at Olive and here in London to train the software to identify segments and allocate them the appropriate place in the hierarchy. The functional possibilities are limited as, because this process works from the bottom-up – i.e. it segments what is on the page rather than working from abstractions – the relationships between levels across the titles becomes confused. Also, the level of detail was felt to be of limited use – do we really need the "Episode" displayed in the hierarchy?

6.0   Our solution

[Slide 18]

Our solution is a more limited version of segmentation in which the hierarchies are present to two levels: departments and items. In the ToC we are only representing the departments: this still reveals the overall structure of a number, while providing an easy to use navigational tool. Because the items are still segmented, they can be subjected to all the content analysis tools we are providing, and will be the basic unit returned as a search result.

7.0   Summary and Conclusion

[slide 19]

This has shaped our editorial strategies for **ncse** in the following ways:

It is a complete an edition of these six titles possible:
- Our *Northern Star*, as far as we know is now the most complete in the world (Kew stuff, multiples, negotiating the portraits).
- And our *Monthly Repository* has all its bits (show *MR* family tree again).
- Our *Leader* has its multiples.

Decided on our target users:
- Academics, of which our key user groups will be specialists in C19th print, and others won't.
- Targeted outreach to identifiable users such as schools
- More general users, who will have an interest in the period, but whose interaction with the material is harder to predict.

And a large part of editing has been research: different user groups require different types of edition; different tools have different levels of accuracy, require greater editorial intervention.
- The content – using indices, searches, text mining.
- But we are not just an archive: the form is represented though Olive's viewer, full page view, segmentation (with metadata), ToC to show structure.

In editing this material today, we, like our nineteenth-century predecessors, are both editors, publishers and hawkers. When considering the periodical as genre, figures such as the author do not offer a suitable organizing principle; rather, the complex relationships that underpin periodical publication force editors to attend to the object as a whole. Whereas authors are conventionally responsible for their words, editors – both of nineteenth-century periodicals and digital editions – must make those words into an object. This involves identifying content, organizing it, developing presentational tools, identifying users, and then making it available to them. As always, this must be achieved under the constraints of time and money. Editing is always located at this intersection between an existing object and a new object: negotiating the relationship between them is not simply about reproduction, but also about transformation. What is important is to remember that such transformations can entail loss, and a fundamental part of the editorial process is identifying where they occur. In the digital domain, the transformation between media means we not only edit the source material, but also the means of making it into something new.