

## ncse Newsletter June 2006

Welcome to the third **ncse** newsletter. Since our last edition, in December 2005, the project team has made progress in the following areas:

### Activities and Publicity:

- Held our first annual symposium.
- Been accepted to present at a number of conferences around the world.
- Re-launched the **ncse** website with substantial updates.

### Research:

- Explored further the multiple editions of the *Northern Star* and *Leader*.
- Continued to develop pilots in collaboration with Olive Software.
- Researched and constructed two alternative models for **ncse**.
- Undertook quality assurance on over 55,000 tiff images.
- Initiated investigations into the potential of keyword extraction and data-mining techniques for metadata population and concept mapping.
- Identified, filmed and digitized additional material to include in the edition



**Read on to learn more....**

## Activities and Publicity

### *Going Digital: ncse 25 February 2006*

In February **ncse** held its first annual symposia. The event featured nine specially invited speakers, from a range of institutional backgrounds. The papers were provocative and challenging, and the audience relished the opportunity to respond. The speakers were:

#### Introduction:

Laurel Brake (**ncse**).

#### Session One: ncse:

Jim Mussell and Suzanne Paylor (**ncse**).  
Harold Short (CCH and **ncse**).

#### Session Two: Going Digital:

Karen Spärck Jones (University of Cambridge)  
Peter Mandler (University of Cambridge)  
Andrew Cutting (London Metropolitan University)

#### Session Three: User Perspectives

Linda Kaye (British Universities Film and Video Council).  
Beverley Kemp (Women's Library).  
Graeme Easdown (Keele University).

#### Respondent:

Marilyn Deegan (CCH and **ncse**).

Laurel Brake's introduction set the tone of the day, introducing the audience to **ncse** and to some intellectual issues that came out of it but were common to all working in the digital domain. Jim Mussell and Suzanne Paylor then built on this introduction, outlining the project in more detail and presenting one possible solution to the increased pagination resulting from the inclusion of multiple editions. Harold Short complemented this paper with an overview of the work of CCH, and some comments about technical solutions that could be implemented within **ncse**. The session was then opened to the audience, and a useful discussion ensued about the project, and its proposed data structures.

The next session was just as lively. Karen Spärck Jones's provocative paper fed back her findings from the recent *E-resources for Research in the Humanities and Social Sciences - A British Academy Policy Review*, and considered what this suggested for the

presentation and organization of data. Peter Mandler's paper discussed the methodological principles that informed the selection of titles within the British Library's *British Library Newspapers 1800-1900*, and reflected on the wider implications of subscription-only services. Andrew Cutting's presentation made an intellectual case for a more imaginative approach to user interfaces: his focus on the importance of games as constituting knowledge provided a counterpoint to presentational methods derived from libraries and archives.

After lunch the final session was focused towards user needs. Linda Kaye's fascinating account of the "user vortex," with users getting gradually sucked into a resource, supplied a dynamic model of user interaction that challenged the more linear models of data retrieval. Beverley Kemp discussed the diverse needs of users, even within the demarcated interests of those coming to the Women's Library. The disparity between user needs and what is available was a salutary reminder to remember the user in the creation of digital editions. Graeme Easdown's contribution covered the various strategies aimed at getting digital resources used within schools. His powerful arguments in favour of these resources, contrasted with the limited success in getting them used, challenged authors to design resources with school users in mind and teachers to develop a pedagogy that embraces such interactive learning experiences.

Marilyn Deegan's able commentary reflected on the day, and drew from the diverse contributions a list of points to consider for **ncse**. The injunction to remember the user has had a powerful influence on our subsequent decision-making, particularly on the appropriateness of the "core" model to different user groups. In all, the day was a memorable one: technical challenges were overcome; refreshments provided (sometimes through ingenious means – our thanks once more to Café Deco who provided coffee and transport at short notice); and useful conversations went on into the evening.

A full program of the event, with downloadable versions of some of the papers, is available here:

<http://www.ncse.kcl.ac.uk/activities/event.html>

### **Conferences and Seminars:**

Over the coming months members of the **ncse** team will be speaking on project-related research at a number of international conferences.

Jim Mussell and Suzanne Paylor will shortly be leading a workshop on both the use of images in the nineteenth-century periodical press and how they might be republished digitally today at "The Verbal and the Visual in Nineteenth-Century Culture" to be held at the Institute of English Studies, London, 23- 24 June 2006.

[<http://www.bbk.ac.uk/english/19c/>]

Jim will also present a paper, co-authored by Suzanne, at this year's BAVS conference [<http://www.bavsuk.org>], at the University of Liverpool, 7-9 September 2006, on "Conflict and Competition in Nineteenth-Century Advertising" which will draw upon **ncse** research.

Laurel Brake has recently given papers in London and Montpellier on "Print Culture in Victorian London" and "Digital work on the Nineteenth-century British Press," and will soon be presenting papers on "Higher Journalism and Popular Culture," "Multiple Editions and the Public Sphere, 1838-52: the Leader and the Northern Star," "Town and Country? Bibliography and multiple editions of two 19C newspapers" in London and New York.

Also, **ncse** team members are presenting papers on their work outside the **ncse** project at various symposia over the summer and early 2007. Full details of these engagements can be found at:

<http://www.ncse.kcl.ac.uk/activities/conferences.html>

### **Outreach:**

As a result of the **ncse** symposia, *Going Digital*, the research team have developed outreach contacts with secondary schools through Graeme Easdown. Together, we are exploring ways in which we might trail **ncse** in secondary schools, perhaps through the Training and Development Agency (TDA), the British Educational Communications and Technology agency (BECTa) or the History Teacher Education Network (HTeN). We hope to get feedback on teacher and pupil experiences of using these materials, to develop exemplar tasks and activities to model the ways in which the resource might support learning across the key stages, and to ensure to the widest possible distribution of the resource through sites like Curriculum Online [<http://www.curriculumonline.gov.uk>].

### **Website:**

In May the **ncse** website was updated with news of our upcoming conference papers and with downloadable papers and presentations from our recent symposium. The website was also recently launched via a number of academic mailing lists and this generated all manner of enquiries and useful feedback from fellow scholars in the field.

\*

\*

\*

## **Research**

### Interesting Fact of the (half) Year!

The Leader is divided into two main portions. The first, set in three columns, provides “News of the Week,” the leading articles, and the correspondence columns “Open Council.” The second, which is set in two columns, was known as the “Portfolio” and contained reviews and fiction under the editorship of George Henry Lewes. In the number for 18 September 1852, Lewes opens his “Literature” column with:

Novelty is vital to a newspaper. If you have not *news* to feed subscribers with, they naturally enough throw you aside: you cease to exhibit your *raison d’etre*, as the philosophers say. And yet when there is no news? Readers generally put up very well with what is called a “flat season;” they are in no dearth of books; and the absence of novelties only enables them to live in more familiar intimacy with the books of a past age.

The news, as its very name suggests, is predicated on novelty, on events that deviate from the norm. As the novelty of news is fleeting – it ceases to become novel, or is surpassed by something newer – newspapers themselves are ephemeral, attached to a moment that passes. Against this Lewes positions literature: unlike news, the “books of a past age” have timeless qualities that permit them to co-exist with the changing lives of readers. As Lewes’s “Literature” column reviews new books, his remarks are appropriate. The publication of these books – and hence his notice of them – is newsworthy, but the criteria on which he judges them are those of literature, and so should transcend the moment of writing. As the first department in the more literary section of a weekly newspaper, Lewes’s review sections mediate between the passing events of news and the timeless qualities of literature.

On the 14 September 1852 the Duke of Wellington died. The subsequent number of the *Leader* – the one that opens with Lewes’s remarks above – records that the “overpowering event of the week, is the sudden, and – to the impatience of a nation’s sorrow – even premature removal of Wellington from the busy world of his conflicts, his counsels, and his glories.” These columns, on the first page of the title below the masthead, are surrounded by a thick black border to represent the loss to the nation (see figure 1. right.). The death of Wellington – an event that “had the effect of surprise” – was the main news story of the week but is not mentioned by Lewes in his



Figure 1. Front page of the *Leader*, 3, 18 September 1852.

“Portfolio.” In fact, this event makes Lewes’s whimsical reflections on news and literature almost tasteless in their tone. What they also do is to foreground the different modes of production of the two sections. The front half of the *Leader* is devoted to the passing events of news and a “Postscript” inserted at the end of “News of the Week” and before the leading articles is dated the day before publication, thereby marking the most timely portion of the number. The literary section that follows contains no dated matter: dealing with newsworthy material (publication of new books, performances on the London stage etc.), but according to timeless criteria, this section did not need to be produced during the week prior to publication. The absence of any reference to Wellington’s death in the “Portfolio” both indicates that it was produced before the preceding Tuesday, and was not deemed worth altering to reflect the momentous event that “all England with one mighty heart mourns.”

### Multiple editions and Selecting the Core:

Since our last newsletter the project team have further researched the multiple editions within **ncse**. We estimated that each number of the *Leader* had two editions, and each of the *Northern Star* up to four. However, the actual figures are less than this: of the 786 distinct numbers of the *Northern Star*, we actually have 2212 numbers, which is an average of 2.8 numbers per week; for the *Leader* the average (taken from the first 3 years) is 1.5 numbers a week, which gives an approximate total of 860 numbers for the 557 distinct numbers in the run. However, this still adds substantially to the pagination of the edition as a whole: we currently estimate **ncse** will incorporate **c98,565 pp**, of which:

<i>English Woman’s Journal</i> (1858-1864):	c6098 pp
<i>Monthly Repository</i> (1806-1838):	c24938 pp
<i>Leader</i> (1850-1860):	c25172 pp
<i>Tomahawk</i> (1867-1870):	c2846 pp
<i>Northern Star</i> (1838-1852):	c18409 pp
<i>Publishers’ Circular</i> (1880-1890):	c21102 pp

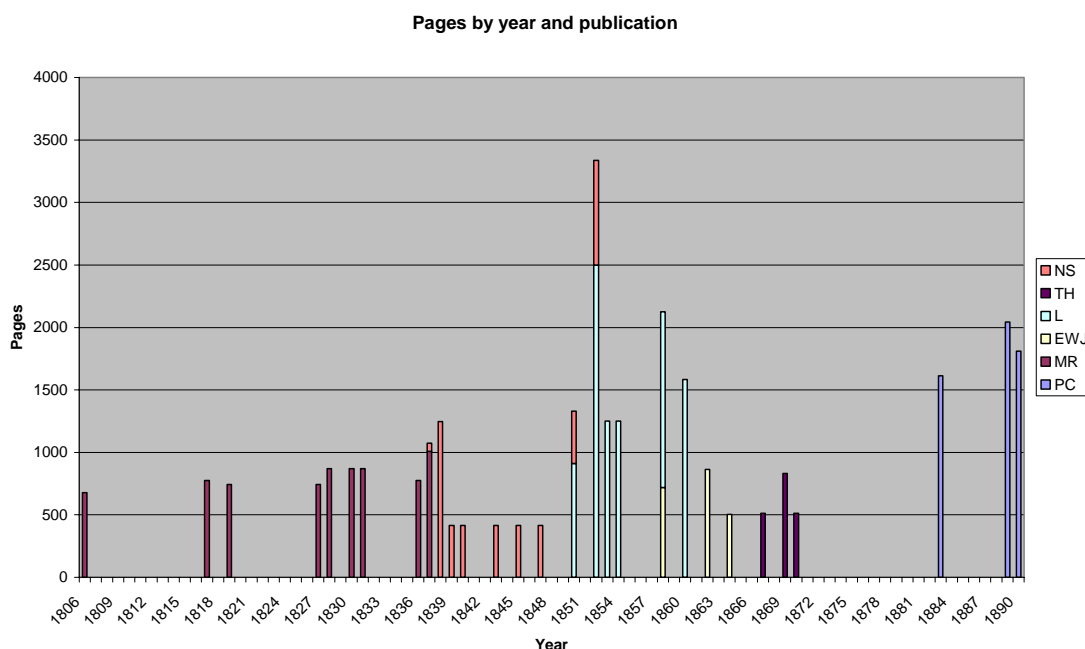
As outlined in our last newsletter, to handle all this material we settled on five principles of selection to delimit a core of 30,000pp that would receive further processing:

- Beginnings: the first calendar year of a title.
- Endings: the final calendar year of a title.
- Changes of Editor: any calendar year featuring a change of editor.
- Visual Material: any calendar year containing substantial visual material
- Thematic Case Study, “Imprisonment and the Press.”

The principles of selection were conceived to provide pages according to categories derived from, and so common to, the genre itself. “Beginnings, “Endings” and “Changes of Editor” select pages where interesting textual variation occur and allow appropriate comparisons to be made between titles. ”The inclusion of “Visual Material” addresses

the textual bias of OCR technology, and ensures that any images can be given the necessary metadata so that they interact with the rest of the content. The thematic case study addresses the under representation of the *Northern Star* and *Monthly Repository* under these principles. Fergus O'Connor, the proprietor of the *Northern Star*, was imprisoned due to a libel published on the 20 July 1839 within a reprinted speech by Bronterre O'Brien. Equally, Robert Aspland, the editor of the *Monthly Repository*, was an active supporter of both William Hone and Richard Carlile throughout their respective prosecutions by the state in 1817 and 1819. By including these years from both these titles, we ensure that pages representing similar thematic content will also gain further processing while addressing the certain shortcomings in the other four principles.

These principles produced the following distribution of years within the core:



As you can see, although the core is not intended to represent chronological continuity, it still contains a reasonable spread across the century. The sparse coverage before 1838 is because the *Monthly Repository* is the only one of our titles then published. The same is true of the *Publishers Circular* at the end of the century: our run of 1880-1890 is the only representative title from the end of the *Tomahawk* in 1870 onwards. There is an inevitable concentration in the 1850s where the *Northern Star*, *Leader* and *English Woman's Journal* at times overlap.

Although we are aware that some users may find the lack of historical continuity frustrating, it should be remembered that this core simply represents the years which will undergo further processing and that the earlier launch of the total edition with minimum processing will enable users to search **ncse** in its entirety.



Full details about the multiple editions, pagination, principles of selection, and the core model can be found in our paper here:

[http://www.ncse.kcl.ac.uk/redist/pdf/06.04.06\\_eventpaper\\_website.pdf](http://www.ncse.kcl.ac.uk/redist/pdf/06.04.06_eventpaper_website.pdf)

### **Pilots:**

Over recent months the **ncse** team has continued to refine the pilots in collaboration with our partners at Olive. The key issue has been how to represent the complex and often irregular structures within the serials through different approaches to segmentation. The processing techniques developed over iterations of the pilots have been able to segment content to a much deeper level than has been achieved by previous projects, and this represents a significant advance in the potential to digitize this material. However because of the often haphazard arrangement of the elements which make up the titles, and the lack of clear visual indicators to mark out the different levels and segments on the page, the team have faced considerable difficulty in developing consistent rule-sets to instruct the software how to deal with the hierarchical relationships between segments. As a result the team are now considering the option of abandoning some of the deeper level segmentation for the immediate purposes of **ncse**, and instead adopting a segmentation model which would, with input and correction from the research team, represent the content of each number at two levels – the ‘department’ and the ‘item’ level. Olive’s team are now preparing another pilot for **ncse** to illustrate what this approach would yield.

### **Models:**

Our original conception of the segmentation was to item level:

Edition (**ncse**) > title (eg. *Leader*) > volume (e.g. vol. 1) > number (no. 1, 30 March 1850) > department (“News of the Week”) > item (“Parliament”)

However, because there are further hierarchies (for instance “Parliament” has a range of sub-sections), it is difficult to automatically determine how these are structured at this deep level across the diverse content of our six titles. We currently have two options: the first is to pursue the two tier model (a processed core and an unprocessed version of the whole edition), within which we would attempt to pursue these structures within the 30,000pp of the core; the second is to utilize Olive software to segment the whole edition – all 98,565pp! – to two levels, with no automatic differentiation between items and sub-items. Each model has its own advantages and disadvantages. The two tier model allows us to demarcate a portion of the edition, and fully recognize all the hierarchical levels of data within it; however, it only allows us to process a small amount of the total, and there will be limited functionality on the whole. If the whole edition was processed, we would have to impose a restricted segmentation onto the periodicals, and then attempt to compensate for this with automatically populated metadata; but we would gain a



consistent and functional level of processing across all six periodicals in their entirety. The decision is a difficult one as on one hand we wish to do justice to the complexities of nineteenth-century print, but on the other hand we recognize that it may not be possible to achieve this with the resources available to us. Ultimately we must decide whether to privilege a core, or to work on a limited conception of the whole; whether we should produce an exemplary edition of a small selection of pages, or attempt to find ways to reproduce structure across the complete runs.

#### **QA:**

Since the arrival of the c.55,000 TIFF images in January 2006 the team have been engaged in an in-depth QA process to identify multiple editions in the *Northern Star* and *Leader*, duplicate frames present in the microfilms which need to be extracted in the digital edition, missing and damaged material which will need to be replaced or re-digitised where possible, to locate visual material, any reproduction problems and to track page sequences and any irregularities in these and validate and flesh-out allocated folder trees with the appropriate bibliographical information in order to construct the user navigation menus in the edition.

#### **Keyword extraction and data mining:**

On the 20 February 2006 the **ncse** research and technical teams met with Paul Rayson, the Director of UCREL [<http://www.comp.lancs.ac.uk/computing/research/ucrel/>] to discuss the ways in which **ncse** might exploit their work on keyword extraction and data mining. Paul introduced the project to the team's work on the CLAWS part of speech tagger [<http://www.comp.lancs.ac.uk/ucrel/claws/>] which may be able to help **ncse** in identifying and extracting indexes of people and places from our OCR text and tagging instances of these with metadata. We also explored the possibility of using the UCREL Semantic Analysis System (USAS) to extract keywords from our OCR text and how the results of this might assist us in concept mapping.

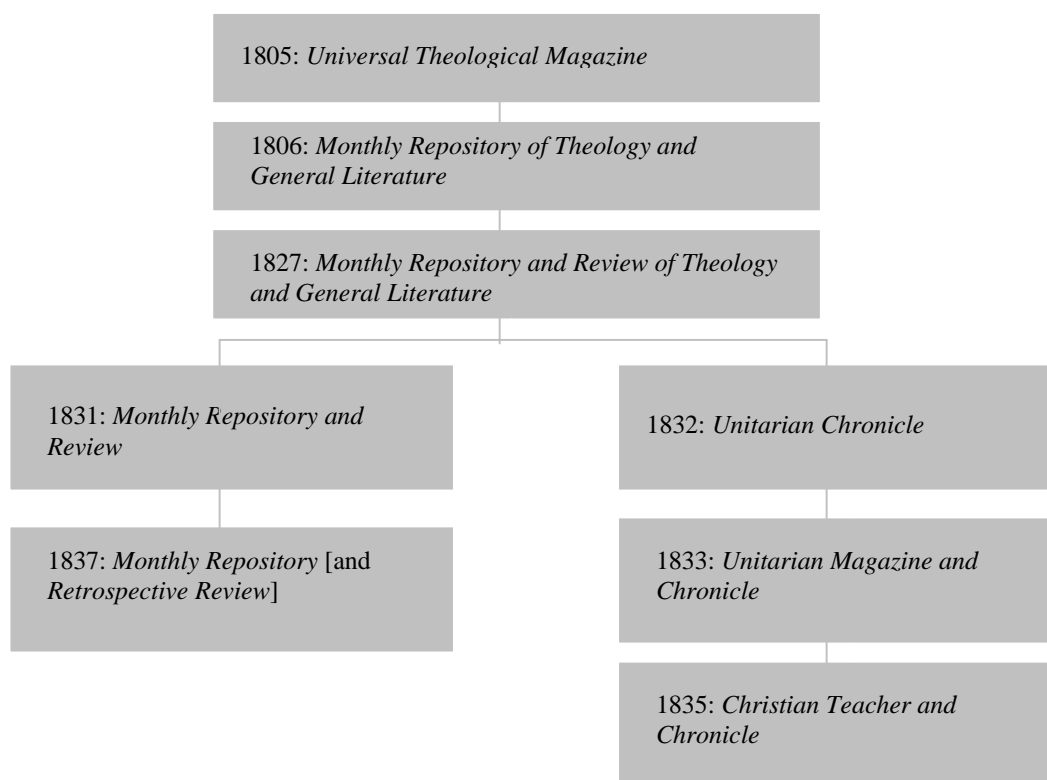
#### **Additional material:**

As part of the page turning process **ncse**'s research team identified a supplement to the *Monthly Repository*, the *Unitarian Chronicle*, which was issued between February 1832 and December 1833 when it became an independent title in its own right. At the close of 1831 the editor of the *Monthly Repository*, William Johnson Fox, bought the title from the British and Foreign Unitarian Association. Fox claimed that the two principles of the *Monthly Repository* – “to advocate the principles, and to record the proceedings, of Unitarian Christians” – were both increasingly popular, and so advocated the *Unitarian Chronicle* as a vehicle for the latter. The opening address in the *Chronicle* insists:

The *Unitarian Chronicle* will also be ‘Companion to the *Monthly Repository*,’ uniform in its size, form, type and paper, so as to preserve uniformity of general appearance. All who wish may therefore bind them up together.

The *Chronicle* was a success, publishing double numbers in June and July 1832 before doubling in size and price in August. However, it seems that Fox’s interest in the title waned during the year – perhaps due to his own troubled relationship with the wider Unitarian community – and it became effectively independent (although still published by him) from April 1833. In January 1834 it became the *Unitarian Magazine and Chronicle* and in January 1836 merged with the *Christian Teacher and Chronicle*.

As **ncse** is an edition of the six titles, we recognize the continuity between the *Monthly Repository* and its supplement:



As with all serial publications, it is difficult to decide when a title becomes something else. In the case of the *Unitarian Chronicle*, we decided to include those numbers where the link to the *Repository* is signalled in the title. Fortunately, the years 1832 and 1833 are bound in a single volume and our partners at the British Library and Olive Software have generously agreed to film and digitize it in its entirety.

\*

\*

\*

## Future work

Within the next six months the **ncse** team will undertake the following:

- **Segmentation.**

The project team will shortly evaluate all pilot material to make a decision between the two models. If we pursue the two-tier model of a processed core of 30,000pp and an unprocessed version of 98,565pp, then we will design segmentation policies and begin work on the early launch of the 98,565pp edition. If we pursue double-level segmentation across 98,565pp, then we will begin work on segmentation policies and begin to investigate the appropriate metadata structures.

- **Early Launches**

If we pursue the two-tier model then we intend to launch phase one of **ncse**, an unsegmented but fully searchable edition of 98,565pp later this year. If we pursue the other model, then we intend to still launch some material, but perhaps to a more limited group as we will continue to work on the material.

- **Conferences and Seminars.**

Members of the project team will be presenting at various conferences and seminars around the world. See the website for details.

- **Carry out user testing events.**

Regardless of the model selected, we will be developing user panels and carrying out user testing in order to design the edition according to user needs.

- **Begin to design a user interface.**

Olive software have begun to design a user interface for their application. If we pursue the two-tier model, then we will also have to work further on the **ncse** website in order to integrate this resource into the existing platform.

**If you would like any further information, or wish to contact the project team,  
please visit our website:**

**[www.ncse.kcl.ac.uk](http://www.ncse.kcl.ac.uk)**

Jim Mussell and Suzanne Paylor

Nineteenth-Century Serials Edition (**ncse**)  
Faculty of Continuing Education  
Birkbeck College  
26 Russell Square  
London  
WC1B 5DQ



Arts & Humanities  
Research Council



**KING'S**  
*College*  
**LONDON**  
University of London



**THE**  
**WOMEN'S**  
**LIBRARY**  
*celebrating and recording women's lives*