ncse: Digitising Journalism

Dr Suzanne Paylor and Dr James Mussell
ncse [www.ncse.kcl.ac.uk]
s.paylor@bbk.ac.uk, j.mussell@bbk.ac.uk
Project deliverables:

- Full-text digital edition of six nineteenth-century serials freely available via the web
- Users navigate facsimile images of pages
- Searching text strings on OCR text
- Searching by bibliographic, generic and thematic metadata
- Contextual and scholarly materials

**ncse titles:**

- *Monthly Repository* (1806-1838)
- *Publishers’ Circular* (1880-1890)
- *Tomahawk* (1867-1870)
- *English Woman’s Journal* (1858-1864)
- *Leader* (1850-1859)
- *Northern Star* (1838-1852)
Research objectives:

- Compilation of most complete edition possible from public repositories and private collections
- Innovative approach to digital editing
  - New model of scholarly editing in the digital domain
  - Implementation of model through new approaches to digitisation and processing and development of hybrid software and bespoke interface
  - New methods of text processing via text mining and data analysis
  - New collaborative working model

Rich page content:

(Text)

An Advertising page from the Publishers’ Circular (1880s)

The position on Fergus O’Connor’s weekly address to readers signals it as important content.
Rich page content:

(Pictures)

Left to right:

- a printed finger post from the *Monthly Repository* in 1814 and a hand-drawn finger post from a margin in the same volume; a finger post from an advert in the *Publishers’ Circular* in 1880 (“all orders”); and the heading to “Open Council,” the correspondence department from the *Leader*.

Fold out illustrations in the *Tomahawk*

Advertising pages from the *Publisher’s Circular*
Mastheads from the six ncse titles and a British Museum Stamp for the back of an illustration plate in the *Tomahawk*
Rich page content:

(Dynamism of form)

Changes of format and paper size in the *Northern Star* over its run.
Structure:
Folder Tree

The ‘edition>title>vol>number’ part of structural hierarchy of the edition is represented in the folder tree.
Structure:
Segmentation mark-up

Extracts from the segmentation policies of the *Publisher’s Circular* and the *Northern Star*
Structure:
The table of components

The ‘Table of Components’ showing OCR of department headers and items as ‘untitled articles’ These will be filtered out before release.
Demo showing snippets in the search results of HTML viewer. We propose to use snippets in the Table of Components instead of uncorrected OCR of department headers.
Structure:
Department>item hierarchy

The English Woman’s Journal and Tomahawk which do not have a clear department>item structure
Editorial issues:
(re)locating material

Pages from the Tomahawk Almanac 1869