

ncse Newsletter December 2006

Welcome to the fourth **ncse** newsletter. Since our last edition, in June 2006, the project team has made progress in the following areas:

Activities and Publicity:

- Continued to give papers at events in a range of subject areas.
- Maintained and updated the **ncse** website.

Research:

- Made substantial decisions about the content and structure of **ncse**.
- Obtained 13 Chartist portraits from the National Portrait Gallery for inclusion with the *Northern Star*.
- Completed a quality assurance check of all 98,565 pages.
- Worked with Olive Software to produce working demos that show the implementation of segmentation for samples from our first titles.
- Designed the **ncse** metadata schema
- Entered the processing stage of **ncse**. Working with Olive Software on a title-by-title basis, we agree a segmentation policy, check the binding of pages into pdfs, and then review the implementation of these across entire runs.
- Continued to investigate the potential of keyword extraction and data-mining techniques for metadata population and concept mapping.



“Christmas Drawings for the Human Race,”
Publishers' Circular, 52, 6 December 1889
(1889), p. illustration 106.

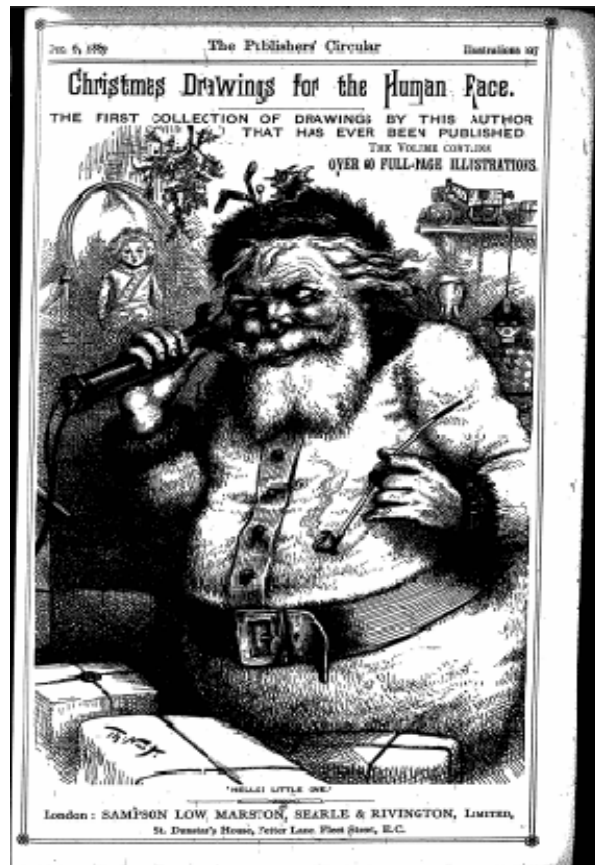
* * *

Turn over to find out who she is talking to....

Activities and Publicity

The Project Team have been busy presenting papers about **ncse** at various events in Britain and the US. Many of our papers and presentation can be found on our website [<http://www.ncse.kcl.ac.uk/activities/conferences.html>].

- “The Verbal and the Visual in Nineteenth-Century Culture.” A two day conference at the Institute for English Studies at the University of London 23-24 June 2006. **ncse** presented a workshop entitled “A picture or a thousand words? The use of images in the nineteenth-century periodical press, and how they are reproduced today.”
- “Text editing, scholarship, print, and the digital world.” A seminar convened by Marilyn Deegan and Kathryn Sutherland at the Centre for Computing in the Humanities at King’s College London, 29 June 2006. Our paper was called “Editions and archives: textual editing and the Nineteenth-Century Serials Edition” and explored the differences between archives and editions in the digital realm.
- “Historical Text Mining Workshop.” An AHRC Methods Network Workshop hosted by Paul Rayson and Dawn Archer at the University of Lancaster, 20-21 July 2006. Our presentation explored ways in which text mining might be used to locate and identify content in archives of newspapers.
- British Association for Victorian Studies Annual Conference, University of Liverpool, 7-9 September 2006. The project presented a paper looking at the advertising pages of the cluster entitled “Conflict and Competition in Nineteenth-Century Advertising.”



It's Santa! “Christmas Drawings for the Human Race,” *Publishers' Circular*, 52, 6 December 1889 (1889), p. illustration 106.

- Research Society for Victorian Periodical Annual Conference, City University New York, 14-15 September 2006. Laurel Brake presented a paper, “Multiple Editions and the Public Sphere, 1838-52: The *Leader* and the *Northern Star*” that addressed the significance of multiple editions both within our edition and nineteenth-century print culture more widely.
- “Publishing Periodicals: Seminars in Book History and Bibliography,” the Institute of English Studies in the School of Advanced Study at the University of London, 23 October 2006. Laurel Brake presented an extended discussion of multiple editions entitled “Town and Country?: the *Northern Star*, the *Leader* and multiple editions.”

Forthcoming presentations include:

- Jim Mussell and Suzanne Paylor will also speak at “Publishing Periodicals: Seminars in Book History and Bibliography” on the 29 January 2007 at the Institute of English Studies in the School of Advanced Study at the University of London. Their paper is entitled “‘A Picture or a Thousand Words’: the use of images in the nineteenth-century periodical press and how they are reproduced today” and is an extended version of the workshop paper presented in June.
- The project has put together a panel for the forthcoming biennial conference of the Society for Textual Scholarship to be held at New York University, 14-17 March 2007. The panel is entitled “Editing Journalism: the Past in the Present” and brings **nese** together with another editing project, John Stokes’s and Mark Turner’s volume of Wilde’s collected journalism for the Oxford English Texts *Complete Works of Oscar Wilde*.
- And of course, the forthcoming **nese** symposium, “Digitizing Journalism,” at CCH on 24 February 2007 will present another forum for the discussion of all things **nese**.



*

*

*

This dispute goes on for over a month, with lively insults and allegations being exchanged in the pages of the *Northern Star*. When Ottley inserts an advertisement under a column of Halifax news that suggests the Tea Dealers remain anonymous only because they are illiterate, the Tea Dealers respond on the front page of the next number of the *Star* claiming that on the evidence of Ottley's prose he must have gained his education from Wackford Squeers – a character familiar to readers of the *Northern Star* as it had extracted the appropriate episode from *Nicholas Nickleby* the previous week. When Ottley seeks to establish his credentials with a grand procession of 113 tea chests through the streets of Halifax, the Tea Dealers cast doubts on his status as honourable tradesman, suggesting instead that he is a leisured gentleman with a horse, dog and gun. The subsequent week Ottley writes that “we have a dog, and pay eight shillings a year tax, for the rat catcher, a *Terrier*, and not a broken-haired *Scotch*; and horses, not for hunting, but for use in our Trade, and that we have not fired a gun this seven years. It is certainly true that we have license to shoot, not with guns, but with *E-pistles* in *Scotch preserves* wherever they may be; and, depend upon it, it shall be renewed from time to time, so long as we have *powder* left and *Scotch vermin* to kill.” Although the puns are not very subtle, they make lively reading and are guaranteed to provoke the Tea Dealers some more. In this way, the debate is prolonged week by week, and the weekly periodicity means that instalments can sometimes be printed side by side on the *Northern Star* front page (above).



Front Page, *Northern Star*, 1, 5 May 1838 (1838), p.1

The dramatic nature of this dispute, played out weekly in the *Northern Star*, blurs the boundaries between news content and advertisement. Indeed, when the first advertisement for Ottley's, “The Northern Locust and his Dupes,” appears the *Northern*

Star's correspondent from Halifax notices the advertisement in his weekly news column, warning that if “*half* the allegations contained in it be true, it is, indeed, high time for the “Dupes” to “Purchase not only Ottley and Co’s Teas, but those of other dealers to see who supplies the best.” The following week, when the Scotch Tea Dealers reply, the same correspondent writes “Audi alteram partem, hear both sides, is a very old and very popular adage” and once again directs readers to the advertisements on the front page in order that readers can do just that.

We are still not sure what exactly occurred beyond the pages of the *Northern Star* to provoke such a public spat. It is likely, however, that the controversy was mutually beneficial for both sides: when compared to the other advertisements, the gossip and potentially libellous allegations of the tea merchants makes exciting reading, especially for a local readership that would know those involved. It is possible that the dispute was manufactured as an elaborate advertising gimmick: a guide to puffing published in 1856 called *A.B. has returned, or the romance of advertising with tracings from the capital of the second column of The Times* suggests just this; and both parties were employing equally devious advertising strategies. However, what is interesting is the way in which the advertising interacts with the seriality of the journal to create a narrative that is itself predicated on news. Such intertextual borrowings are common in periodicals. As serial texts, their link with time ensures an engagement with news. News, in turn, announces new things, and these might be products as well as events. This overlap allows advertisers to link their products to novel events or, as in the case of Ottley, pretend their advertisements are really notices for the public benefit, much like news items themselves.

Nowhere is this more apparent than in the activities of the publisher of the *Northern Star*, Joshua Hobson. In the *Northern Star* for 31 March 1838, Hobson reveals the circulation of the *Northern Star* in Yorkshire is on average 10,659 weekly. This information was political at the time as the *Star's* main rival, the Whig *Leeds Mercury*, was selling just under 10,000 copies a week. Immediately below this notice is an announcement that the *Star* is printed on a new 2 h.p. steam engine, with an endorsement for Messrs. Smith, Bencecock and Tannett, engineers of Leeds. This is not the only time when Hobson combines radical politics with commercial activity. In March 1842 an advertisement for “Farr’s Life Pills” is disguised as news following a report from the House of Lords with the announcement of “Great Excitement” (above right). In the same number Hobson, who clearly relies on the ready incomes that patent medicine advertising provides, decides there is more money to be made by competing with them. His advertisement for

HOUSE OF LORDS.—MONDAY, MARCH 14.

Lord BROUGHAM gave notice of his intention to move a series of resolutions respecting the proposed income-tax. The purport of these resolutions, which he read at length, is condemnatory of the imposition of an income-tax, unless under very extraordinary circumstances, and also affirming certain principles respecting the apportionment of such a tax, with relation to the distribution of capital and income amongst the various classes of the British community.

Earl FITZWILLIAM made a few observations on the subject; and after the Lord Chancellor had laid on the table of the House a Bill for the amendment of the bankruptcy law, the House adjourned.

GREAT EXCITEMENT.—The discovery of Old Parr’s secret has caused great excitement among the members of the medical profession, as it seems probable, ultimately, to exterminate a good deal of their contradictory nonsense. Many members of the profession have, however, honourably declared their intention of using Parr’s Life Pills, after witnessing their astonishing curative properties and re-invigorating powers on the human frame.

“A New Theory,” *Northern Star*, 5, 14
March 1842 (1842), p. 6.

“Chartist Pills” makes clear that these pills will “avert much of the illness usually affecting the Working Classes” and “give the suffering Chartists an opportunity [...] to forward the cause of Democracy, and assist the families of their incarcerated brethren” (right). The pills can be obtained from either Hobson or the Proprietor, Fergus O’Connor, and are only available from the *Northern Star* offices. As well as being publisher and printer of the *Northern Star*, Hobson is here also a commercial advertiser. However, this latter role is not that different from his more familiar roles. After all, the entire letterpress could be conceived as an advertisement for the *Northern Star*, encouraging readers to buy it again next week.

Such activity reminds those of us involved in digitization projects that the categories we use are often insufficient to capture the complexity of both text and the various historical actors that created it. The process of transforming historical artefacts into a useful digital form entails re-inserting certain categories of information back into the material. For instance, structural divisions that are indicated on the page by lines must be encoded in order to reproduce the divisions between types of content; equally, the activities of historical actors like Hobson that are indicated in the text also demand attention in order to connect the traces with information about individuals. The *Northern Star* tea debates demonstrate just how difficult it is to distinguish between news and advertisement, or between printer, publisher, author, or advertiser. The complexity of the text demands sophisticated acts of labelling: the amount of work required to do this reminds us exactly how much information is packed into the print on each page.

Designing the Edition: the core

At the time of publication of our last newsletter in June 2006 the Project Team were in the midst of deciding between two different models for **ncse**:

CHARTIST PILLS.

IMPORTANT TO THE AFFLICTED.

MR. J. HOBSON, *Northern Star* Office, Leeds, having accepted the Wholesale and Retail Agency of these Pills, is authorised to give Two-pence out of each 1s. 1½d Box, to be divided between the Executive and the Families of the Imprisoned Chartists.

The many Medicines lately offered to the public would have prevented the proprietor from advertising these Pills (although convinced of their efficacy), did he not feel it his duty to give his suffering fellow Chartists an opportunity (by their affliction) to forward the cause of Democracy, and assist the families of their incarcerated brethren.

To no one is health so important as to the Working Man, when deprived of it his means of subsistence are suspended and his sufferings aggravated by reflections on his poverty, and the helplessness of his family.

These Pills are not put forth as a cure for all diseases, but their use will avert much of the illness usually affecting the Working Classes. The Proprietor has witnessed with pleasure their extraordinary efficacy in numerous instances of loss of appetite, head ache, heart burn, palpitations of the heart, bilious and nervous diseases, pains in the stomach, and other symptoms indicating an imperfect action of the Digestive Organs.

To those of sedentary habits whose trades confine them in an unwholesome atmosphere, and perhaps for hours together in a continued posture, thereby inducing habitual costiveness, indigestion, and nervous debility, they are strongly recommended and have been found of essential service, as they enable the system to throw off those morbid accumulations which occasion disease, at the same time strengthening and giving tone to the stomach, and invigorating the whole system, by these means establishing health on a firm foundation.

Until Agents are appointed generally, those persons who wish to try them can receive a box, with ample directions, post free, by sending One Shilling in silver, and two Postage Stamps, in a pre-paid Letter, to Mr. J. Hobson, *Star Office*, Leeds.

Applications for Agency will be attended to, and the terms sent by return of Post, by writing either to Mr. Hobson, or to the Proprietor. All Letters to be Pre-paid.

“Chartist Pills,” *Northern Star*, *Northern Star*, 5, 14 March 1842 (1842), p. 2.

1. a two-tier model, with a fully-segmented core of 30,000 pages complemented by an edition of the whole of **ncse** 98,565 pages with minimal processing.
2. a single-tier model in which all 98,565 pages are processed uniformly to two levels of segmentation: departments and items.

There were persuasive arguments for each. The two-tier model allowed us to more accurately represent the textual hierarchies present in the periodicals, but due to our limited resources not over the whole edition. The single-level model would produce a more user-friendly edition with all pages being processed to the same specification, but necessarily at the cost of the more sophisticated segmentation needed to capture all of the relationships on the page. After many discussions both internally with the project team and with members of our advisory board at our event in February, we eventually recognized that a complete, uniform edition would be of more use to users than one that only offered full segmentation on a frustratingly sparse range of pages. As a result, we opted for option 2.

As we were developing what option 2 would look like with Olive, we learnt that it was only possible to implement segmentation to one level. This meant whereas we had intended to segment both departments and items:

- Edition > Title > Volume > Number > Department > Item

We could only segment:

- Edition > Title > Volume > Number > Department

Or,

- Edition > Title > Volume > Number > Item

Segmentation is the process that identifies textual units on the page and, as these in turn drive much of the functionality of the edition (they are returned in search results, they are units that can be printed etc), it was clear to us that the latter option was preferable. However, the decision to apply segmentation over 98,565 pages rather than a core of 30,000 pages meant we had to consider carefully the consequences for our workload. For

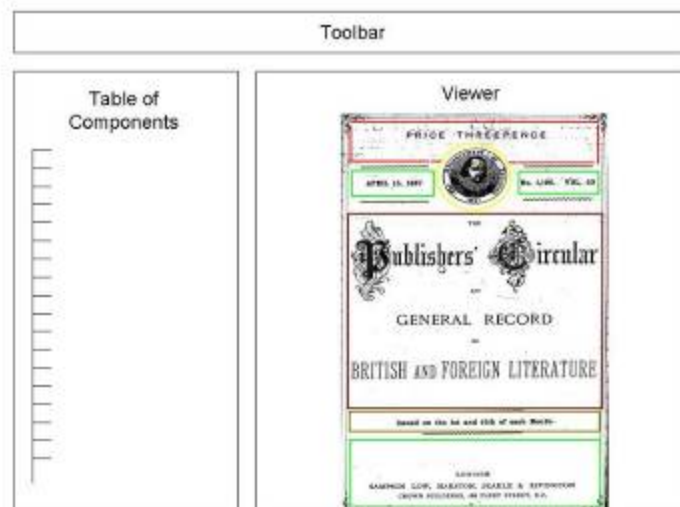


Diagram of layout of Olive Viewpoint viewer for ncse.

instance, in each number of the *Leader* there are roughly seven departments but something like 147 items; this translates to 7028 departments over the whole run, but 147,588 items! Should we choose to add metadata to each textual unit – for instance to give them all titles or identify them as a certain type of content – it would clearly be a great deal easier to only do so on departments. That said, items remain the most coherent textual unit: it is with individual articles, rather than sections of articles, that people want to work.

The decision was also complicated by the delivery mechanism that is part of the application that Olive we developing for us. As you can see from the diagram on the right, on the left of the screen is a “Table of Components” that lists the contents of each number and allows the user to get an overview of where they are while they browse the facsimile pages on the right. As the contents of this Table of Components is populated from whatever we determine are segments, it would contain either a list of departments or items. A list of departments would provide an easy-to-read survey of the types of article in each number, and allow users to click and jump to sections which interest them. A list of items, which would also include things like department headers because they are also components on the page, would be much harder to navigate because of its length and, as the text itself is derived from our OCR, it would need to be corrected extensively, and by hand.

So we were left with another difficult decision: item level segmentation seemed to offer the best functionality at the level of the page, but choosing it would necessitate much more work in the Table of Components; department level segmentation was the best solution for the Table of Components, but did not really work in terms of page content. Our decision was to go for a hybrid solution. We would opt for item level segmentation but then edit the contents of the Table of Components so that it would only show the department headings. This would allow users to see the contents of each number, according to the labels on the pages themselves, while reducing the amount of hand correction to a more manageable amount. Working with Olive, we even devised a way in which everything but a department heading would have the same label in the Table of Components, allowing them to be easily identified and suppressed.

Processing ncse

After a substantial period of planning and negotiation, we have finally begun production of the **ncse** materials. We are working closely with Olive on a title-by-title basis to complete the following tasks:

- Design a segmentation profile. This is a set of rules that will enable the Olive system to identify all the items in a title. We are also trying to find rules that can distinguish between department headers and the rest of the content, so that we can keep the hand editing of the Table of Components to a minimum.

- Feedback on demos. Olive test our segmentation profile on a few sample numbers of each title in order to examine how the rules have been applied and to signal any exceptions.
- Complete quality assurance. Olive have processed the original preservation tiff images, and bound them into pdfs that correspond to individual serial numbers. This process is complicated by the presence of multiple editions, wrappers, volume frontispieces and supplements, so we go through page by page and check the numbers are complete.
- Check and edit the segmented material. After processing has been completed, we sample numbers from the run in order to check that the segmentation profile has been correctly applied. We also work to correct the Table of Components.
- Complete the folder tree. The hierarchy of content in **ncse** (see above p.8) is represented through a folder tree. It is important that the right files and folders are in the right places, and each is labelled appropriately.

These tasks are very time-consuming, and each title throws up unexpected problems. At the time of writing we have completed the segmentation profiles of the first two titles (*Leader* and *Publishers' Circular*) and are about to begin the third (*Northern Star*). Processing has begun in Israel, and we are expecting the *Leader* to be segmented imminently. We are working to quite a tight schedule, as we aim to complete the processing of all six titles by May 2007.

ncse metadata

There will still be plenty to do in May, however. One of the consequences of opting for a more simple segmentation policy over the whole 98,565 pages rather than a more complex segmentation policy over just a core of 30,000 pages is that any interventions we make must be applied to three times as many pages. This is especially problematic for the level of segmentation we have lost – department level – as this needs to be recreated using metadata on every page. The Olive application we are using to deliver the edition does support metadata, but only non-relational, simple text strings that must be entered by hand. As we do not have the resources to insert metadata by hand for 98,565 pages – and certainly not for all the items on each page – we are working with Olive to develop the means to machine process as much metadata insertion as possible.

We have already had some success with this. As our segmentation policy does not recognize departments:

Edition > Title > Volume > Number > Item

Items will not be associated with the section to which they belong. So, in the *Leader* the segmentation policy would recognize:

ncse > *Leader* > 1 > no. 2, 6 April 1850 > Lord John Russell in Manchester

without linking “Lord John Russell in Manchester” to its department, “News of the Week.” As the departments are listed in the Table of Components, and these entries correspond to items such as headers on the page, we have devised a system in which the department label is cascaded down to all items that occur between one department header and the next. This means every component on the page between “News of the Week” and the next department header, “Public Affairs,” is marked with metadata that links it to the appropriate department. Although not entirely accurate (for instance, there are often items that stand apart from departments), this reinserts a level of hierarchy that would otherwise be lost.

The full **ncse** schema is on the next page.

New Olive Product: Viewpoint

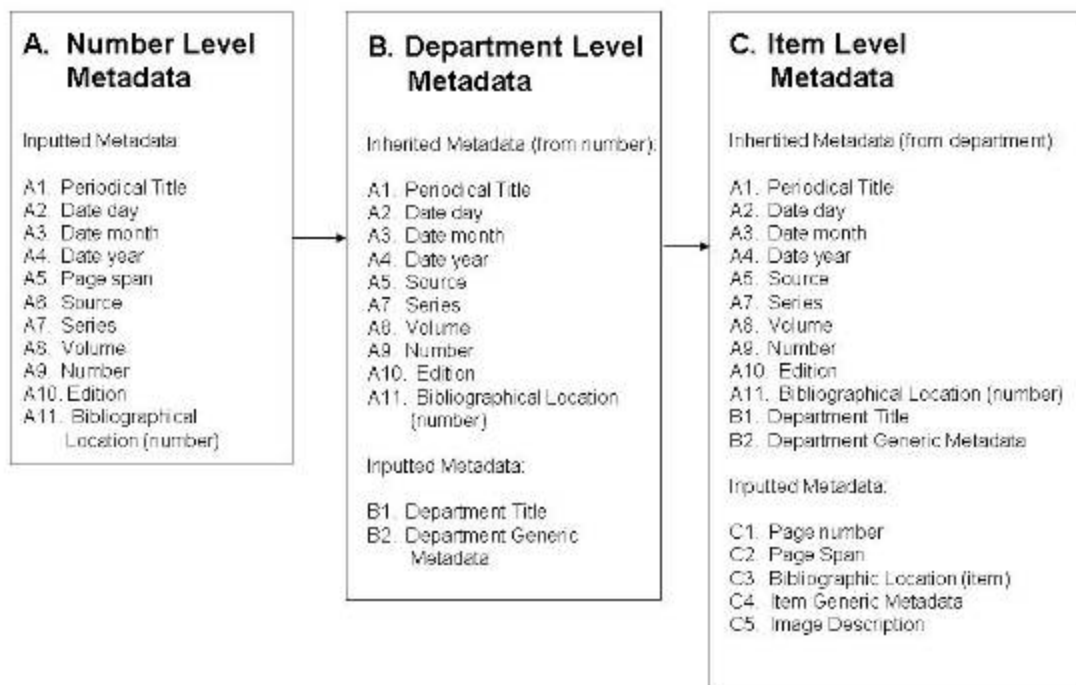
One of the interesting outcomes of our initial experiments in digitizing nineteenth-century periodicals was that their diversity challenged our existing ways of categorizing print forms. As you can see from the image at the end of p. 3, our titles look very different: some, like the *English Woman’s Journal* are quite book-like while others, like the *Northern Star* look like newspapers. Olive have two existing applications, Active Paper and Enterprise Publisher: the former is designed for digitizing newspapers and the latter for a wider range of content, but premised on the model of the book. We found that different titles worked better in different applications, but there were features that we wanted to retain from both. Olive have been developing a new application, Viewpoint, which can handle a wide range of material while allowing customers to select which components make up the final application. **ncse** is functioning as a beta test project for the development of Viewpoint, and we are pleased to have the opportunity to lend our expertise to the design of a product that can be of use to others in the future.



THREE SPEECHLESS SOUTHERNERS IN THEIR ROCKING CHAIRS.

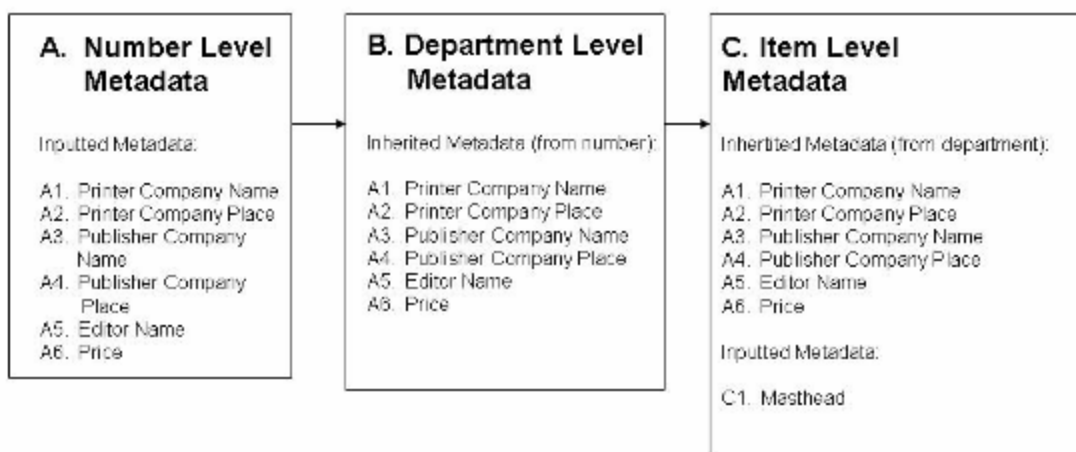
Detail from “Vizetelly and Co’s Recent Publications,” *Publishers’ Circular*, 45, 6 December 1882 (1882), p. illustrations 81.

ncse Essential Metadata Fields



There are also metadata categories that we would like to add, resources permitting. These are:

ncse Extra Metadata Fields



*

*

*

Future work

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

- **More processing**

The processing phase is scheduled to last into May. We are working through the titles in the following order:

1. *Leader* (25,172pp)
2. *Publishers' Circular* (21,022pp)
3. *Northern Star* (18,409pp)
4. *Tomahawk* (2,846pp)
5. *Monthly Repository* (24,938pp)
6. *English Woman's Journal* (6,098pp)

- **Text Mining**

The decision to segment all 98,565 pages has increased the amount of material we have available to experiment with text mining. At present our investigations are mainly focused on deriving keywords from items that could indicate both their subject and genre. We will also work with extant lists of data to index the edition.

- **ncse symposium: Digitizing Journalism**

The second **ncse** symposium is scheduled for 24 February 2007 at CCH, King's College London. At the time of writing the program is still being finalized, but there will be presentations and demonstrations from **ncse**, and then papers from invited speakers including Hilary Fraser (Birkbeck College), Ed King (British Library) and Jerome McGann (University of Virginia). Places are limited and there will be a small charge for lunch. If you would like to attend, please contact the project team.

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



“Messrs Jarrold and Son,” *Publishers' Circular*, 51, 6 December 1888 (1888), p. 1523.

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

