

“ncse: Publishing the Nineteenth-Century Periodical Today”

I would like to begin, as so often in discussions of periodical literature, with a quotation from W.T. Stead. In his annual *Index to the Periodical Literature of the World*, Stead writes, and I quote, “Periodicals are by their nature the most evanescent productions of the press. The world of periodicals is perpetually changing. Old periodicals are passing away; new periodicals are being born. Some continue to live under the old name but with entirely different tables of contents. In other cases the old contents reappear under a new title. Changes of price, of address, of circulation are constantly taking place. In short, all the phenomena of growth are conspicuous everywhere in the world of periodical literature.” Stead’s metaphor here grounds the dynamic nature of the periodical press in the embodied, organic world. Indeed, in another volume of his *Index* he claims, and I quote, “In the world of magazines there is every year a copious list of births and deaths. Marriages are rare, although of occasional occurrence [...]” Here Stead not only imagines the mortality of periodicals in very human terms, but even sanitizes the copulation and reproduction of periodicals, implying both their sex lives and their gender, through the metaphor of marriage.

Stead’s periodicals, like the anthropomorphic figures that run down the edge of our fliers, seem to operate in a self-sustaining world in which the products of the press are free to move, exchange and reproduce themselves. This, of course, is a conceit – the periodical press is structured by elaborate networks of publishers, contributors, editors and distributors, not to mention readers, or the various material components and transformations that underpin production – but it is an effective way to capture

an often unstable field. Writing annually in his *Index*, Stead is not surveying the changes in the press as they happen, but instead is framing his analysis from the different perspective provided by his annual periodicity. This is not navigating a flow of publications as they appear, as his monthly *Review of Reviews* does, but rather sorting and indexing a fixed archive, within which changes have occurred. Stead's periodical people are not running the race like those on our fliers, but rather pausing to be counted.

Stead's anthropomorphic periodicals are markedly different to a more recent survey of the field. Jerome McGann, as part of his intervention into *Critical Inquiry's* recent assessment of contemporary critical theory – and recently published in *Interdisciplinary Science Reviews* – advocates “hands-on collaborative work” as the means to learn the skills to participate in the digital revolution. He writes, and I quote, “The library, especially the research library, is a cornerstone if not the very foundation of the modern humanities. It is undergoing right now a complete digital transformation. In the coming decades – the process has already begun – the entirety of our cultural inheritance will be transformed and re-edited in digital forms [...] Theoretical as well as very practical discussions about these matters have been going on for years and decisions are taken every day. Yet digital illiteracy puts many of us on the margin of conversations and actions that affect the centre of our cultural interests (as scholars and educators).” If the library is going digital, McGann argues, scholars should have a stake in how this occurs: not only are our resources being reinterpreted on our behalf, but our career structures depend on a form of paper-based publication that is becoming increasingly obsolete in the wider commercial marketplace. However, the rhetoric of digital publication often tends

toward the utopic: “Digital scholarship” McGann warns “even the best of it – is all more or less atomized, growing like so many Topsyies. Worse, these creatures are idiosyncratically designed and so can’t easily talk to each other. They also typically get born into poverty – even the best-funded ones.” Like Stead’s periodicals, digital editions must be nurtured, taught how to communicate, and sustained by institutional support. For McGann these are immediate, pragmatic concerns, requiring an urgent shift of attention from textual analysis and critical theory informed by the logic of paper, to a recognition of the different demands of a digital publishing format.

It is between these two axes that I want to introduce you to the Nineteenth-Century Serials Edition, **ncse** for short. The project, now eight months into its three year span, will produce a digital edition of six very different periodical titles, complete (where possible) with advertisements, prefaces, and supplements. In order to navigate this diverse material, we are embedding metadata to allow searching within a range of delimiting fields, and also implementing concept mapping, to facilitate research within configurations of conceptual and thematic categories. The six journals are taken from across the nineteenth century, and represent the diversity of forms and periodicities within the genre: the *Monthly Repository*: a metropolitan magazine, at first Unitarian but increasingly non-sectarian, that contains literary commentary, biblical criticism, and significant early nineteenth-century poetry; the *Northern Star*: a heady mix of mid-century working-class radical political commentary, international, national and local news, poetry, entertaining tit-bits which was, of course, the organ of the Chartist movement; the *Leader* – about which Laurel will speak in more depth later – a progressive, liberal and secular weekly

established by Thornton Lee Hunt and George Henry Lewes; the *English Woman's Journal*: a mid-century, middle class monthly that embodied the feminist goals of the Langham Place group; the *Tomahawk*: a satirical alternative to *Punch* that contained fold-out comic illustrations; and finally the *Publishers' Circular*: a long-running, London-centred, fortnightly trade publication for the print and publishing industry. Taken as a cluster, the journals challenge our conception of the periodical itself. Whilst the *English Woman's Journal*, an unillustrated monthly containing essay reviews, poetry, correspondence, and commentary, fits readily into our notions of the nineteenth-century periodical, the *Tomahawk* is an illustrated weekly that self-consciously spoofs the periodical as genre with its mock reviews, essays, news, leaders and adverts. The *Northern Star* meanwhile is self-consciously a newspaper; but the *Leader*, whilst also considering itself a newspaper, has a more ambiguous identity, containing two discrete sections, one concerning news and current events, and a second, positioned at the back, containing literary commentary and essay reviews.

Publishing these six titles as a single cluster alters their status: where they might otherwise be considered complete titles within their own historical cultural contexts, or alternatively repositories for types of content that participate in scholarly narratives elsewhere, by publishing them together we have foregrounded both their generic similarities as periodicals, and also the shared cultural references that are inscribed despite radically diverse contents. The cluster thus foregrounds not only the interconnectedness of print culture, but also the serendipity inherent in textual research: in an abundance of possible contexts, our circumscribed set of six journals is delimited in ways that further provide countless objects, whether images, marks,

words, articles, departments or pages, that can be reconfigured, searched for, and, most importantly of all, read.

This politics of abundance is also inscribed both in the paper-based periodical archive and the potential offered by developing digital technologies. The size of the archive – John North estimates some 125,000 titles – is further complicated by its formal diversity, range of content, and unique relationship to time. Equally, the hyperlinked pages of the world wide web offer a proliferating network of contexts; while digital technologies offer the potential for the categorization, identification and juxtaposition of texts and objects in unexpected ways. As such, Stead and McGann, in different ways, provide an important frame for the objectives of **ncse**. Stead surveys a paper archive, delimiting and demarcating it in ways that make it manageable while still recognizing its mutability and dynamism. McGann argues for a hands-on understanding of the demands of digital publication, in which digital objects behave very differently from their paper-based analogues. Stead, even though he is distanced by his annual perspective, is surveying his contemporaries; McGann, although also surveying the work of his contemporaries, addresses the demands of reinterpreting historical works from the past via their surviving remains in the present. **ncse** recognizes the complexity of the periodical as genre, archive and artefact, and applies technological solutions in order to capture and make accessible the forms and contents of these six periodicals, for analysis, and wider comparisons in the digital contexts of the present.

*

*

*

Whereas both Stead and McGann reach for organic metaphors to describe their respective modes of publication, of course neither is actually dealing with living things at all. In addition, their respective objects behave very differently. The success of paper is largely due to its portability, reproducibility and (relative) stability over time. The lack of interoperability between digital resources is the direct result of their means of production: although markup languages such as xml and standards such as the TEI have gone some way to standardizing practice, metadata standards, that would define data in ways that are machine readable and so could be located and manipulated across resources, are still being developed; and the requisite institutional frameworks – whether in terms of technical support, legal advice, collaborative networks, or archival spaces – are a long way from realization. For Stead, these problems do not exist in the same fashion: his periodicals can happily get married as they are in a form that fosters amalgamation and juxtaposition while participating in a wider institutional economy that actively encourages such activity. McGann, citing Swinburne, writes, and I quote, “Today some new things have to be mastered. In addition [...] we have to be concerned about these new things, about how we make them and what we use them for. We will do this by becoming students again – a role that, as educators and humanists, I think we’re especially good at.” Although this is sensible and timely, the rhetoric of mastery seems uneasy. Although we are indeed responsible for the objects and their meanings that circulate in the world, they too affect us – whether in terms of ecological damage, the influence of the market, or the role of commodities in the structuring of desire. Equally, although I work on a digital project, I rarely – if at all – feel like the master of my laptop. Although McGann’s rhetoric of mastery is related to his wish to operate at the level of production and use rather than remain in

what he sees as an over-theorized superstructure, it seems that this distinction between theory and practice is rather arbitrary. Editing is a form of intervention and so implies a politics: rather than eschew theory in an embrace of practice – to adopt a “citizen’s life” in McGann’s words – it seems we should re-apply our critical resources to our paper archives, digital tools, and resulting electronic editions. Digital resources are subject to historical factors and so cannot offer definitive versions, or function as surrogates for paper archives. They are instead alternate versions, and so we must interrogate the continuities and similarities between the digital and paper versions of these historical artefacts.

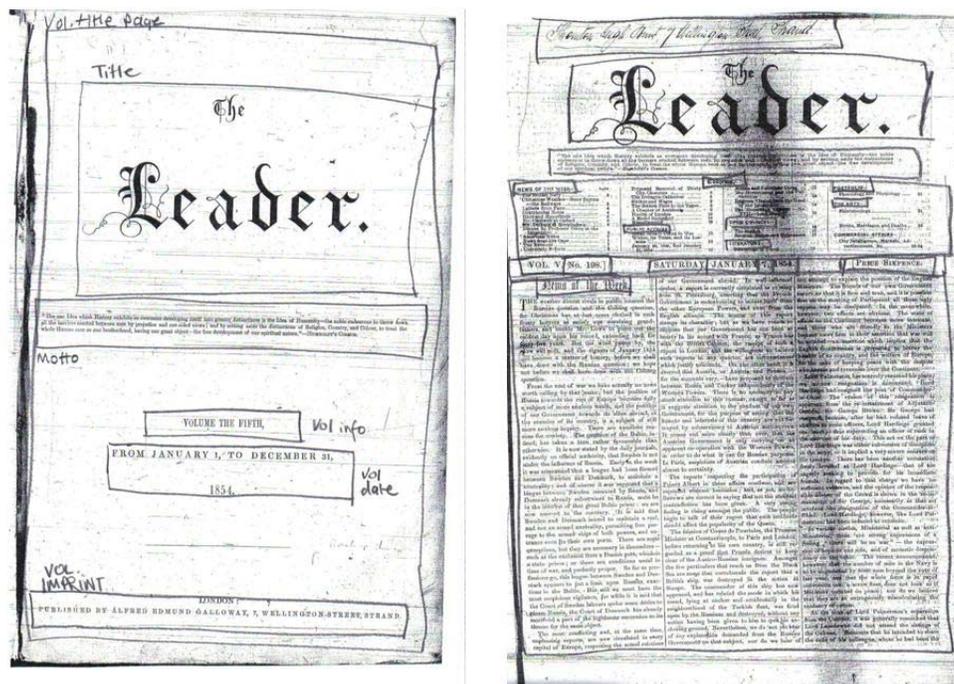
It is this policy that we are attempting to embody in **ncse**. It is well recognized that books and journals are not just repositories of important texts, but rather they constitute the meanings of those texts. By offering facsimiles of the periodical pages, complemented by a searchable, automatically generated full-text, we can maintain, and make available formal features that might otherwise be lost. E-texts such as Project Gutenberg, while providing valuable access to otherwise scarce or expensive works, are not adequate for historical analysis as they efface their means of production by providing “pure” text. In relying on this transcribed text, the semiotics of typography, as well the spatial politics of the page, are lost. However, as we insist on the importance of features such as headings, images, dividers, advertisements, and imprints – as well as refuse to over-privilege the run, volume, number, item, or text string – we have radically increased the corpus of material within **ncse**. Although the project team contains specialists in the nineteenth-century press, we recognize that the limits, and indeed biases, of our expertise, coupled with the scope of the edition, render any attempt at mastery necessarily

partial. What we seek to do is to empower readers to encounter whatever it is they are looking for within the context of the periodical. Rather than extract content from the abundance of print, we want to ensure that content is easily located but remains situated within it. To do this we have developed four levels of markup with which to structure the edition and so enable users to configure it according to their own interests. These are: segmentation; structural metadata; advanced metadata categories; and concept mapping. In facilitating the simultaneous use of these four levels within the user interface, we intend to open the full complexity of periodical publication for our users.

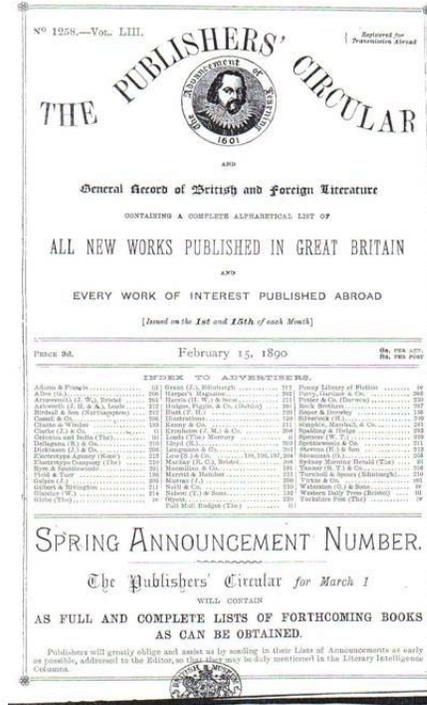
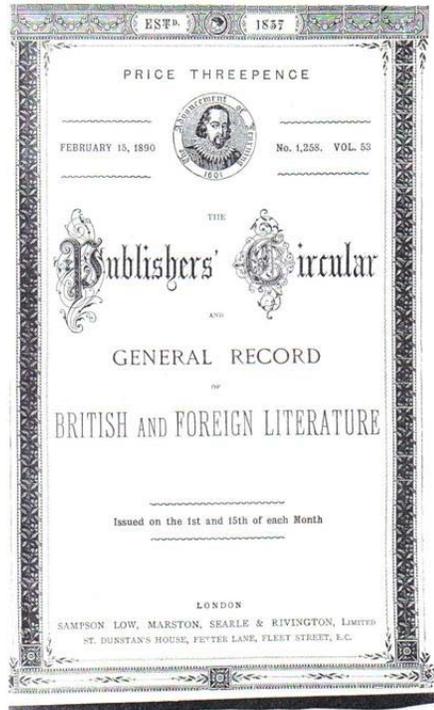
The simplest way to recover data from the edition will be free text searching from the automatically generated text. However, the processes through which the text is generated from the microfilm source is not 100% accurate, and the size of the archive limits the extent to which we can correct it. By using fuzzy searching the software, to some extent, compensates for the inevitably inaccurate transcription. However, for some of our titles, this is not sufficient. The *Northern Star* for example, being a cheap and rather flimsy working-man's weekly, is in poor condition for the application of the technology that produces the textual transcript. It was poorly printed, and smudges of the impression, tears, and creases often obscure the letterpress. In addition, the poor quality paper on which it was printed has deteriorated markedly, and so the edges of pages and quality of the letterpress have damaged further.

It is the privileging of text over form, however, that represents the most significant limit of free-text searching, no matter how sophisticated. Unable to identify images,

some typographic fonts, or spatial layout, such searches do not directly reflect what is on the page – in fact they explicitly ignore it – and so take little account of the contextual references that structure meaning. The first level of markup, segmentation, directly addresses this problem. The research team, in collaboration with Olive and our partners at the Centre for Computing in the Humanities at Kings College London, are developing ways in which we can use imaging technology to recognize and demarcate formal features on the page. Preliminary trials have suggested something like this will be possible:



These features are often structured in quite complex hierarchies. For instance the title of *Publishers' Circular* is not the same on the cover of the wrapper as it is on the first page of the paginated letterpress:



This necessitates the occurrence of “title” as a subset of both wrapper cover and front page. When we remember that the *Monthly Repository*, which ran from 1806–1838 and is in four distinct series, also has a different title on its various volumes as well the first page of each number, capturing the variants in a structure that applies to all titles results in a complex data map.

It is unlikely that we will achieve all of the units demarcated on the data map: it is designed to be exhaustive rather than realistic, and the categories on the left are easier to capture than those categories towards the right. However, it is essential that all the units that we can demarcate through the imaging process and its subsequent refinements, are located within the wider context of the cluster. The second level of markup, structural metadata, is information that is added to segmented units in order for our search engine to identify, sort, and delimit the appropriate digital objects. Again, the scale of the edition limits what we can

achieve in these fields. It is likely that this metadata will have to be entered automatically either as part of the imaging process, or through automated scripts. Either way, the data will be sufficient to identify an object in terms of its location within the edition, and, crucially, allow the juxtaposition of the units that we segment.

The third and fourth levels of markup exploit the code that is produced by the imaging process in order to provide more interpretative means of navigation. At the third level of markup, advanced metadata categories map indexical terms throughout the edition, allowing users to trace their occurrence across and within the titles. We have identified seven main categories: people, places, events, institutions, products, publications and businesses. Although we recognize the overlap between these categories – a publication can be a product, and a person can be a business – they are designed to delimit data in order that it can further be explored. We are currently working to construct a database, built from extant lists of data, to contain this information in a way that we can manipulate to overcome some of the limits of free-text searching. For instance, much of the nineteenth-century press is pseudonymous: by marking up all people, it becomes possible to link authors to their pseudonyms, perhaps uncovering the more surreptitious activities of authors and editors in their texts. Equally, by identifying place names, we can situate them within a hierarchy – for instance identifying the “Strand” as a place important for the print trade, but also as a district within a larger entity called “London.” These categories also allow us to engage with indirect and non-textual material. When the “Great Exhibition” was being constructed it was very rarely called by this name. The *Tomahawk* especially refers to the Crystal Palace through a number of subtle

allusions – usually referring to its resemblance to a crystal cake – none of which would be recovered by direct searches. Equally, the cartoons in *Tomahawk* – a key part of the journal taking up, in some cases, a quarter of its pages – would not feature in text searches as even the captions are often wittily oblique.

The fourth – and we've imagined it as the highest – level is concept mapping. This is an interpretive search tool which will operate by identifying thematic content. In a series of discussions we have identified conceptual terms and linked them in a three layer hierarchy. These concept maps will allow users to reconfigure the edition according to their research interests: for instance those interested in the urban will be able to delimit all items that deal with the experience of the urban by selecting the master category environment, and the subset urban. Again, the scope of the edition limits the extent to which we can interpret content and mark up accordingly: although we are exploring key word extraction, lexical databases, and semantic linking, computers cannot read, and so are no substitute for expert interpretation. We are still in the process of developing both the conceptual and electronic tools to implement concept mapping, but the objective is to provide flexible search methodologies that will bind the edition together in unorthodox ways. As the concept map is both evolved from the journals, but not dependent upon them, we hope that it will be both a valuable research tool in their own right, and also – as a product of a specific moment in nineteenth century studies – become a historical artefact. By mapping the edition thematically, we hope to produce a new perspective on this material, complementing historical narratives with an emphasis on the interconnectedness of nineteenth-century print culture.

The combination of these four levels will enable users to navigate the edition while also providing a digital representation that preserves the architecture of the periodical. By providing digital images of the pages as well the facility to perform complex searches on the generated text, the digital edition will render our cluster more accessible than either its nineteenth-century paper form or twenty-first century electronic transcript would alone. Simultaneous searching across the edition will emphasize its nature as cluster and, while we make no special case for the interdependence of these six titles, demonstrate the inevitable cross-references inscribed within it. Indeed, such serendipitous encounters are central to research into the periodical press: the volume of material insists that research will never be exhaustive and, just as before digitization, there will still be an element of chance to what users find. What **ncse** does is to map data fields in a way that opens them up, offering users the chance to reconfigure the edition while retaining the historical integrity of the paper-based archive.

It is only through recognizing the malleability of the periodical archive, just as Stead did in his *Index*, and then incorporating the digital tools to preserve this in its electronic form, that we can remain true to the bibliographic challenges of the press. Periodicals are more open texts than books, employing a wider visual repertoire and inviting a number of reading strategies in their efforts to reach readers. It is thus important to sustain these features in any electronic version, even if this radically increases its complexity. **ncse** recognizes the contextual nature of the press – that it does not exist in isolation at the level of title, department, article, or even word – and inscribes this within its internal structure. As McGann argues, not only do digital technologies offer us an opportunity to explore historical artefacts in new ways, but

there is an increasingly pressing need to engage with their politics. **ncse** has begun to do just this, producing a flexible edition that, if wider projects such as McGann's NINES come into being, can also operate within the equally unexpected contexts of other developing nineteenth century digital archives.