

ncse: Selecting the core

Introduction:

From Google's 'man on the moon initiative' to get all upward of 30 million books past and present online and searchable (Google print) and the rival Open Content Alliance project backed by Microsoft, Yahoo and the Internet Archive, the proliferation of electronic resources looks set to continue to gather momentum. Nineteenth-century materials have been a key growth area here, Google print, for example, will initially make digitise c. 1 million volumes from the Bodleian's nineteenth-century collections.¹ The reason for this predilection for nineteenth-century materials is multifaceted. The unprecedented volume and diversity of material printed in the nineteenth century make it a good source of experimental and diagnostic materials – as well as one which engenders great academic and public interest. However, nineteenth-century printed matter is also free from the constraints and legal entanglements of copyright issues and presents fewer OCR challenges than the typography and irregular and unfamiliar language of previous centuries. Furthermore, an ever increasing proportion of this treasury is also in desperate need to preservation beyond the budgets of holding repositories hampering user access. In recent years there has been a massive upsurge in the number and the scale of projects digitising all manner of nineteenth-century printed matter. The venture has attracted a diverse set of participants from large commercial publishers like Thompson Gale and Proquest, keen to be key players in the market for subscription only services, to JISC, and the British Library and a whole host of smaller-scale BA and AHRC projects like **ncse**. In addition, a whole host of supporting materials are also now becoming available, from indexes like the invaluable Waterloo and SciPer to attempts to integrate and organise electronic resources for scholars, like Jerry McGann's NINES initiative.

¹ <http://infoday.com/newsbreaks/nb041220-2.shtml> [22.02.06]

In this paper we hope to outline the **ncse** project, to place it in this context and to discuss some of the challenges we face in realising **ncse**'s goals. In particular we will focus on the issues involved in selecting material for deeper-level processing. It is partly this processing which, it is hoped, will make the output of **ncse** especially valuable not only as an electronic resource but also as an investigation into how scholars might become more directly involved in the digitisation of humanities sources.

ncse is a three year AHRC project which aims to meet two ends. Firstly to produce a full-text digital edition of six diverse nineteenth-century serials and make this freely available via the web. Optical Character Recognition (OCR) technology will enable users to search the full text of the whole edition whilst viewing readable facsimile images of the pages on which their highlighted search strings appear, or to simply browse the edition wholly or by page, issue, volume or title. Second, **ncse**, aims to formulate and implement new ways of realising our current scholarly conceptualisation of these materials in electronic form, allowing users not only to engage with the structural and textual elements of these objects, but to use those elements to delineate different types of content in searching and browsing. In order to achieve this end **ncse** is engaged in producing models of the material's content, physical features and semasiology. Though they will not be fully realised within **ncse**, these models can then be drawn upon to design the information architecture of the resource and as models for future projects. The resource takes advantage of the latest technical resources to make freely available an electronic archive of significant historical interest in its own right. In addition to this, we are using this archive as the context with which to develop experiment digital tools that substantially redefine it. **ncse** is thus of its moment, archiving historical documents of current scholarly interest, while also contributing to wider developments in humanities computing. In addition we plan to apply advanced processing to a selection of the material to in order to showcase and explore the ways in which his processing can open up these materials to users in unconventional ways.

1. The titles:

In order to maximise the potential of **nese** to fulfil these aims the titles chosen for digitisation are deliberately diverse in terms of their form, content and readership and span across the century. Together they represent some of the diversity and intricacy of the morass of print journalism and its communications networks in the nineteenth-century. This diversity will also pose interesting technical and conceptual challenges in rendering and processing.

Whilst the *English Woman's Journal*, fits relatively readily into the periodical convention the others have been chosen to purposefully stretch the template. The *Tomahawk* for example was a weekly, packed with illustrations and self consciously sets out to spoof the periodical as a genre with its mock reviews, literary essays, news, leaders and even adverts. Similarly, the *Publisher's Circular*, a fortnightly publication, was packed with adverts and functioned as a trade journal to the print and publishing industry. The *Northern Star* meanwhile was self-consciously a newspaper, and the *Leader*, whilst considering itself a newspaper, in terms of content had a more ambiguous identity, being demarked into two clear sections, one concerning news and current events and one containing literary commentary and essay reviews.

The periodicals are also disparate in terms of content and purpose. The provincially-focused *Northern Star* was a heady mix of mid-century working-class radical political commentary, international, national and local news, poetry and entertaining tit-bits, and functioned as the official organ of the Chartist movement. The metropolitan and cosmopolitan *Monthly Repository*: was a radical, middle-class, at first Unitarian and latterly non-sectarian with its literary commentary, poetry, biblical criticism, and reviews. The *Leader* was a progressive, liberal and secular weekly headed by Thornton Lee Hunt and George Henry Lewes and reflecting both scientific and literary interests. The *English Woman's Journal* was a mid-century, feminist yet respectable and middle class publication. The *Tomahawk*: was a satirical alternative to Punch packed with comic illustrations and the *Publishers' Circular* was a long running London-centred trade publication for the print and publishing industry.

In order to digitally represent this motley miscellany of genres, styles, intentions and readerships the team will have to develop novel approaches to conceptualising the material which at once both facilitate its systematisation and meaningful integration, and retain its material and semantic integrity. This work we hope will lead both our

team and our end users to engage with the wider archive of nineteenth century journalism from a fresh perspective and reconfigure the genre ‘periodical’. We hope to achieve some of this by abstracting and ‘concept mapping’ the content of our serials in ways which take account of the physical and referential context in which the content is situated, and extra-textual elements like illustrations and adverts.

As the periodicals represent a variety of forms, they present interesting technical challenges for the team in terms of how we render the material in a consistent manner. The *Northern Star*, for instance with its closely set broadsheet columns of the grubbily printed text, provides a stark contrast to the crisp and well set demy octavo pages of the *English Woman’s Journal* and the lively illustration packed fold-out pages of the *Tomahawk*. The edition will give users a flavour of the wealth of choices available to nineteenth century readers. This multiformity presents interesting technical challenges for the team in terms of how we render the material in a consistent manner. In order to facilitate cross-title searching and comparisons we must conceive of this disparate material as a set of structural units common to all the titles in the edition and which reflect current scholarship. *More on this later*

2. The existence of multiple editions

The unanticipated existence of multiple editions in the hard copy runs of the *Leader* and the *Northern Star* had a huge impact on the project. We had agreed at the outset that we would pursue all identifiable components of each title, including advertising wrappers and supplementary material, and that this principle should apply to multiple editions. However, the number of multiple editions – usually two per number in the *Leader* and anything up to eight per number in the *Northern Star* – substantially increased the number of pages that would be included in the edition as whole. This increase, from an originally projected 30,000 pages to an estimated 110,000 pages, would place a considerable strain on the resources of all those involved in the project, and would necessarily restrict the amount of processing possible before the project end in 2007. In this section of the talk, I want to explain why we thought it was essential that multiple editions should be represented in the edition, and then lead on to how we are managing the increased workload.

The publication of multiple editions was common practice for many nineteenth-century titles, but this varied between periodicities. Many monthly periodicals offered readers the chance to purchase in quarterly or annual units, some offering ready bound editions that, indexed and eliding the space between numbers, resembled books. Just as today's newspapers – both online and paper-based – publish editions throughout the day in order to update the news, so nineteenth-century titles produced images that narrowed the temporal gap between numbers. Weekly periodicals, positioned often uneasily between the daily press and the more reflective monthly review, often combined content that focused on the passing events of news with more self-referential serialized pieces. As such, it was important for weekly periodicals to contain all information up to their date of publication. Even with the proliferation of the railway and its accompanying telegraph network, it still took time to distribute numbers to readers located far from the site of publication. Rather than allow spatial distances between readers to determine how up-to-date a title could be, many weeklies – the *Leader* and the *Northern Star* included – offered town and country editions, with the country edition often published a day before the town edition in order to reach its more distributed readers by the date of publication.

Judging from the hard copy at the British Library's newspaper collections at Colindale, it appears that the *Leader* publishes two editions, a town and a country edition, from its fourth number. The only differences between these two editions is the department called 'Postscript', occurring immediately after the survey of the week's events, 'News of the Week', and before the leading article, 'Public Affairs'; and the financial information printed at the end of the number in 'Commercial Affairs.' The 'Postscript' is inserted after the rest of the number has been printed, and often carries its own date to show how contemporary it is. The country edition of the *Leader*, which of course is the first edition of that week, uses the 'Postscript' from the previous week's town edition. The financial information is usually derived from the *London Gazette* and, by publishing this later in the town edition, they can use a more current number of the *Gazette* as the source. By the eighth number of the *Leader*, the edition is indicated on the first page. The presence of these editions at such an early stage registers the *Leader's* aspirations to be more than a metropolitan weekly, but still privileges the readers in town with more up-to-date information.

The *Northern Star's* use of multiples is much more complex. From its first volume in 1838 the *Northern Star* published up to five numbers for each week, but it is impossible to say from the bibliographic evidence whether five represents an exception, or whether other editions are missing. We know that the Colindale run of the *Northern Star* is incomplete, as in 1840 there is only one edition per number, but these are often marked in the letterpress as second or third edition. However, it appears that the run becomes more consistent in its later years: the run was derived from the Stamp Office and, from 1844, a clerk conveniently records on the front the number of each edition. This is valuable information as the actual textual differences between editions are often slight, sometimes not occurring at all. This raises some perplexing questions: if a number is not marked as a multiple, and has no discernible differences from the other numbers of the same date, is it in fact a duplicate and should we include it in the edition? We have opted to include as much of the Colindale run – itself the basis for most of the microfilm runs of the *Northern Star* currently throughout the world, and the forthcoming run included in the British Library's *British Newspapers 1800-1900* – within the edition as possible, regardless of the extent of textual change, as we hope that our further processing and research will allow us to analyse what differences there are in much more detail.

The policy behind the *Northern Star's* publication of multiple editions is difficult to uncover. While based in Leeds they published a first edition that reached country agents in England and Scotland by Saturday morning. This edition (3,000) was part-printed Wednesday and completed Thursday before presumably being distributed on Friday. They then published at least two other editions on Saturday (37,000), presumably for readers in Leeds and late in the country. After the move to London in 1844, the title published a first edition for country agents – again including Scotland – on Saturday. There are then at least two further editions on Sunday, one for the country and another (the latest they publish) for London. This distributed pattern of publication, with numbers being published between Thursday and Sunday each week, challenges our assumptions about weekly publication. The *Northern Star* is not only a Saturday weekly, but also a Friday and a Sunday weekly, and it not only represents London news, but different constructions of the country too. Under each masthead, in other words, are a number of different numbers, all determined by a distinct combination of spatial and temporal factors.

Although the mastheads of weekly numbers of the *Leader* and the *Northern Star* suggest that there is only one number for each date, and that number is part of a sequence of volumes and numbers stretching backwards and forwards in time, the bibliographical condition of these titles is much more complex. We are very comfortable, both in terms of our intellectual frameworks and archiving practices, with the idea of linear runs of periodicals, principally as they conform to well-established book-like patterns. It is this logic that privileges just one edition, seeking to establish bibliographic order over what it is a much more complex material reality. Rather than reproduce this in the architecture of our digital edition – essentially nominating one edition as a copy text – we feel that digital tools allow us to offer users the chance to specify the material that they work with. Within the terms of its cluster of six titles, **ncse** endeavours to represent the surviving complexity of nineteenth-century print culture, and is committed to developing a framework that renders it navigable.

3. Our solution:

As the decision to include multiples radically changes the scale of the edition, we have had to reconceive it in ways that accommodate the extra pages without sacrificing its original principles. Our solution is a two-stream resource: the first, phase one, is a full repository of all the pages we have identified, including multiple editions, supplements and advertising wrappers where possible; and the second, phase two, is a selection from this material of 30,000 pages that will receive advanced processing. Phase One is due to be published later this year and will be a significant resource of approximately 110,000 pages, with a searchable OCR text and facsimile images of the scanned pages housed within a unique interface currently being developed by Olive. We hope to also give users the option to filter out multiple editions from their search results if they so chose and we are currently exploring methods of automatic processing which may provide further searching options.

In the phase two release of the resource currently scheduled for March 2008, three types of advanced processing will facilitate complex searching of a selected ‘core’ of 30,000 pages of our resource. At the most basic level preliminary segmentation

developed in collaboration with our partners at Olive, will break down the facsimile images into hierarchically arranged navigable units within each title, which, it is anticipated, will include title, volume, issue, department and department items (i.e. article, illustration or advert). Further work to develop this segmentation may then be carried out by the research and CCH technical team. This segmentation is chiefly intended to facilitate precise navigation, but it is hoped might also be developed to allow users to direct text string searches of the full OCR text to specific levels of content. In addition, because the rationale behind this segmentation will be consistent across the edition, users will be able to perform delimited searches simultaneously across the publications.

At the second level, users will, it is hoped, be able to delimit and direct searches using additional structural and advanced metadata fields. The structural metadata fields will distinguish different types of content from one another, e.g. articles from adverts, and are derived from this comprehensive data-mapping of our sources. This frighteningly complex diagram was our attempt to systematise and abstract our conception of the structural units into which our sources could be broken down, and the relationships between them. Whilst its realisation is far beyond the scope of **ncse**, we do hope that it will be of utility in future projects and that it may invite scholars of this material to think about it in new ways.

The advanced metadata fields: people, places, and perhaps even events, will allow users to search across the edition for references, both explicit and oblique to people and places listed in our database. To facilitate such searches we hope to work with existing databases, like John North's extensive Waterloo Directory and to explore the possibility of developing databases of people and place references from our own OCR text using the CLAWS part-of-speech tagger to mark-up occurrences of proper nouns.²

Such advanced metadata categories explicitly address some of the limits to free-text searching. For instance, much of the content of the nineteenth-century press is signed

² <http://www.comp.lancs.ac.uk/ucrel/claws/>

in pseudonym: 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. These categories also allow us to engage with indirect and non-textual material. When the ‘Great Exhibition’ was being constructed and the Crystal Palace built, it was very rarely called by either of these names. The *Leader*, for instance, routinely calls it the ‘Exposition’ and *Tomahawk*, alludes to it though a range of conceits, including a ‘crystal cake’! All of these variants would be lost in 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 often even the captions are wittily oblique. In these ways we hope to use advanced metadata to open up to interrogation material that would be lost in OCR only searches, and so to therefore remedy the over-privileging of textual content in the digital representation of our journals.

At the third and highest level **ncse** will investigate the possibility of incorporating search fields based on our concept-mapping of these sources. This concept map would be used to generate a set of hierarchically related subject classifiers by, which users could delimit searches thematically, selecting a single theme or combinations of themes which map across the whole edition.

These classifiers, developed from the journals themselves, but yet reflecting current scholarly nomenclature, will also, it is hoped, interact with the lower level metadata fields so that users could for example search for ‘the domestic’ only in adverts. The objective of providing searches based on concept map fields would be to provide non-linear search methodologies that would bind the edition together in unorthodox ways. 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 different types of searches would enable users to navigate the edition whilst 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 OCR text, the digital edition will render the cluster more accessible than either its paper forms or OCR plain text would alone. Simultaneous searching across the edition would emphasize its nature as cluster and

(while **ncse** makes no special case for the interdependence of these six titles) demonstrate the inevitable cross-references inscribed within it.

4. The Core:

In order to select the pages that were to receive this further processing the research team came up with a set of principles of selection. In a series of conversations, we discussed the merits of various strategies, before settling on a final set of five. We considered obvious options such as only processing certain titles, sacrificing the cluster in order to treat a run in its entirety. We also considered selecting by date range, allowing a historical period, or perhaps an event, to select which pages were processed. Both of these were rejected as in neither case would the core represent the edition as a whole. As we have mentioned, the six titles are simply one possible selection from a much larger corpus, and the titles that were chosen were selected to exemplify the diverse forms of periodical publication. We felt that similar criteria applied to this further selection: faced with a potentially endless range of possible cores, our final core should be faithful to the larger edition, and allow us to demonstrate features of the genre.

The scale of the periodical archive, at all levels of analysis, ensures that totalizing claims remain elusive. Both exercises – selecting the titles and selecting the core – are necessarily partial selections from a much wider excess of print. The incompleteness of the core is not necessarily a failing, as the gaps that it introduces signal clearly the choices we have made, and so stress certain continuities while gesturing towards the absence of others. Selections based on total runs posit a completeness that they do not fulfil: if we had opted to include a whole run of one title, we would have to be certain that we had all of its supplements, editions, and advertisements. Equally, selecting according to an event would necessitate intensive research to trace all subsequent mentions of that event in the edition. Date ranges too prove troublesome: there is little overlap between our six titles and the serial nature of periodical literature ensures the titles would always look beyond the limits that it prescribes. Instead, we derived principles of selection from the genre that the six titles, in all their diversity, ostensibly share. By selecting these generic features we

make no unsustainable claims as to completeness; rather, we stress certain continuities across titles that do not depend upon sequentiality or an undefined extratextual context.

The principles we have chosen are:

Beginnings

Endings

Changes of Editor

Visual Material

Thematic Case Study: Imprisonment and the Press

The first two categories are slightly problematic in the periodical context, as beginnings and endings are always provisional. For instance, the *Monthly Repository* was created when its editor Robert Aspland bought William Vidler's *Universal Theological Magazine* and changed its name. This name change signals a beginning, but only at the cost of suppressing a wider continuity between these two members of the Unitarian community. Equally, its shifting allegiances over the long run are represented in its changing subtitle: when Aspland was bought out by the Unitarian Association at the end of 1826, they change it from *Monthly Repository of Theology and General Literature* to *Monthly Repository and Review of Theology and General Literature*. When William Fox bought the *Repository* outright from the Association in 1831, he amended this to *Monthly Repository and Review*, dropping the theology just as the title moved towards a more radical, literary and secular orientation. In its final years, under the veteran editor Leigh Hunt, it simply becomes the *Monthly Repository*, although it does gain a supplement called the *Retrospective Review*. Supplements, of course, complicate things further: the *Monthly Repository* produces the *Unitarian Chronicle* in 1832 to contain the Unitarian business Fox has no room for in the journal proper. Its beginnings are located in its difference from the *Monthly Repository*, which continues alongside it. In 1833, however, it severs the connection completely, becoming the *Unitarian Chronicle and Magazine*. This too is a beginning, and the journal carries on under this title until 1835 when it is merged

with the *Christian Teacher and Chronicle*. This ending, of course, is also a beginning, and so this discontinuity embodies a further continuity.

The third principle is designed to produce material that captures editorial interventions into the periodical texts. Although not all editors alter the format of the titles that they edit, it was common for new editors to reconfigure the title in order to attract new readers. By selecting material where a change of editor takes place, we hope to be able to subject these textual variations to further processing. The fourth addresses the technical processes of digitization. Because we are developing a transcript from OCR technology, any visual material (or unconventional typology) will not be captured. By including those pages where visual material appears, we will be able to add metadata that can address the textual bias inscribed in our methodology.

As we worked through the four bibliographic principles, we maintained statistical information that recorded the percentages of each title represented in the core. This information revealed that our principles favoured some titles over others: although the core was not intended to represent the quantities of each page in the whole edition, but rather the diversity included within in it, we felt that we should devise a final principle that could include more pages from under-represented content. To complement the generic principles, our final principle is thematic but still connected to the press. The two under-represented titles, the *Northern Star* and the *Monthly Repository*, are both the longest running titles in the edition and also have a strong interest in radical politics. Proprietors of journals were legally responsible for what was published in their pages, and the reactionary political climate of the early nineteenth century led to a number of brushes with the courts. The proprietor of the *Northern Star*, Fergus O'Connor, was imprisoned in 1840 for a libel printed in the first edition of a number from the previous July; and Robert Aspland, editor of the *Monthly Repository* for its first twenty years, was active in supporting both William Hone and Robert Carlile, both imprisoned for selling seditious pamphlets. By including the years in which these events take place, we complement our emphasis on form with an acknowledgement of the role of content.

To implement the principles we had to stick to a rigid methodology. We decided that a suitable unit to include was a whole a calendar year as, even though this does not

correspond to the volumes, it is a recognizable unit to readers who are not familiar with the titles. We also recognized that the historical utility of the core was compromised by its fragmentary nature. We decided that we would include any single years that were omitted between two years that were included, creating mini-spans within the core. We were also conscious of the temporal overlaps between some of the titles, and as the further processing is designed to facilitate comparisons, we were keen to retain these where possible. However, these further methodological principles were strictly secondary, and the limit of 30,000 pages prevented us from including many more volumes.

The other problem confronting us was what to do with the multiples in the core. As the core is both representative of the edition, and designed to explicate certain generic features of print culture, it is important that the multiples, which after all occupy up to 40,000 pages of the whole, are represented. However, including multiples for all the selected years in the core would severely restrict the number of years we could include. A compromise was reached: we chose to include multiples from the *Northern Star* and *Leader* for 1852 as this is the final year of the *Northern Star*, and the year where the two editions in the *Leader* are indicated on the front pages and well-represented for most numbers; and also the multiples for 1838 of the *Northern Star* as this is its first full year and overlaps with the last numbers of the *Monthly Repository*.

Although our core selection, which totals 30,284 pages, incorporates years from across the century, there remain regrettable gaps. Only the *Monthly Repository* represents the first forty years of the century, and our thematic case study is an attempt to address this by providing additional volumes from this portion of the century. Also, as none of our titles are from the 1880s, this decade is unrepresented. There is a concentration in the middle of the century due to the overlaps between the *Leader*, the *Northern Star* and the *English Woman's Journal*, and the large number of pages in 1852 produced by the presence of multiple editions. This concentration around 1850 corresponds to the lifespans of the initial six titles, and so the core is a fair representation of the temporal bias that is inscribed in the edition as a whole.

The core is principally to showcase the experimental processing that we intend to carry out. Although its incompleteness might restrict its use for data-mining, it should

be remembered that we are providing a complete edition of all the years in addition to this partial selection. As this larger edition includes a sophisticated search engine, it is hoped those users will use this to survey the material and, if what they are looking for is in the core, take advantage of the further tools provided there. For those interested in the press, the pages we have selected will be those of most use, and allow detailed analysis and comparisons across the titles. The principles privilege formal diversity in their focus on moments of instability such as beginnings, endings and changes of editor, and so incorporate the widest range of material within the core. The focus on visual material directly address the logocentric bias of OCR technology, and forces us to create metadata schema to include the visual with our categorization of textual letterpress. Ultimately, the core is a diagnostic space within which we can develop and demonstrate the experimental further processing and tools we outlined earlier; we believe that these principles offer the greatest variety of material on which to work, while making persuasive arguments about nineteenth-century print.

Conclusions:

Serial literature embodies both discontinuity (the space between numbers, changes in personnel) and continuity (the next article, the next number, the next volume). The core participates fully within this logic: by including whole years, and runs of years where possible, we have substantial sequences of numbers that demonstrate immediate continuities. The fragmentary nature of the core inscribes the arbitrary choices that must be made to delimit material in any consideration of the periodical press, while stressing the continuities between titles that are occluded by an over-emphasis on the integrity of the run. This includes the generic aspects that are addressed directly by the principles of selection, as well as the connections that will be exposed as a result of the advanced levels of processing. Advanced metadata, for instance, will allow users to trace not only the occurrences of historical actors across titles, perhaps bringing to light some of their more surreptitious activities, but also any subsequent discussion of them. Digital transformation provides additional interpretive contexts: just as the structural metadata ensures form is comparable across the digital edition, so the concept maps provide the means to navigate through content. In this way the core makes an important argument about print culture: as its

discontinuities – both in pages selected and the types of processing we carry out – produce further continuities, it emphasizes the interconnected nature of nineteenth-century print culture beyond overdetermined conceptual categories such as those derived from the book.

The two streams of **ncse** roughly address its two goals. Phase One functions as an archive, holding for the first time these formerly distributed and fragile objects in one place. Phase Two is a diagnostic space that explores different ways of representing and making accessible serial literature from the past. Despite the great value in providing this material freely online to users, we recognize that as an archive **ncse** is strictly limited. The project is funded by an AHRC Resource Enhancement grant, and it is in this value-added content that many of the merits of the project lie. **ncse** is a project that both acknowledges the specific demands of periodical literature and develops conceptual, digital and archival tools to address them. For instance, the new interface developed by Olive is in direct response to the uncertain textual condition of periodicals: their two existing programs, Active Paper and Enterprise Publisher, are designed to present newspapers and books respectively; periodicals are neither of these, and so they have had to combine features from each. The various advanced processes we are developing with CCH are experiments in extracting meaning that redefine the way periodicals are structured and manipulated. The contextual materials that we are providing, both in terms of metadata and accompanying documents such as headnotes, provide cutting-edge research that is derived from the edition and will shape the way in which it is used. In the current climate of large-scale digitisation projects it is vital that humanities scholars participate fully in order to delimit material in ways that represent both current interpretive frameworks while respecting historical derivation. However, the process of implementing digital projects is one of collaborative imaginings; and it is important to realize that this process is itself research, helping to create the context for subsequent imaginings. At **ncse** the contribution of academics ensures we produce a product that is sensitive to the material it holds, but the realization of this is possible only through the research of all of our partners.