

# Indexing *The Athenaeum*: aims and difficulties

Micheline Hancock-Beaulieu and Susan Holland

The acquisition of the marked file of the Victorian journal *The Athenaeum* led to the current project based at City University, London. Relational database software provided the means to create a unique bibliographic research tool of reviews and reviewers, supported by an authority file with biographical details. Problems of inclusion, identification, attribution and indexing procedures are discussed. How the index will be made available to scholars is briefly considered.

## Introduction

*The Athenaeum* Indexing Project at City University was launched in 1988, with joint funding from the British Library and the University. The periodical has long been a rich source for literary scholars and the need to index it had been voiced in different quarters over the years. However, the project became a reality because of a combination of two important factors: initiative and resources. Firstly, Professor Oscar Wellens of the Free University of Brussels, who was working on the marked file, began to transcribe the necessary data for the makings of the index. He was later assisted by a colleague from the University of Ghent, Dr Marysa DeMoor. Secondly, it was recognized by the information staff at City that the technology for handling this vast amount of data was available and could make the endeavour feasible within a reasonable length of time. Previous indexing projects, such as that of *The Strand Magazine* and *The Wellesley Index*, could not take advantage of the advancements of information technology.<sup>1,2</sup> The latter took more than twenty years to complete.

This paper will present some of the background to the *Athenaeum* Project, the organization, the state of progress and some of the problems encountered so far.

## *The aims of the project*

The aims of the project are twofold:

1. to index the reviews and reviewers of the periodical from the marked file;
2. to create a database which will serve as a research tool providing comprehensive access to the different data.

## *The Athenaeum and the marked file*

*The Athenaeum* is a literary and scientific weekly periodical which ran from 1828 to 1923 (when it became the *New Statesman*). Not only did it have such a sustained run, it was and is well regarded for its broad review coverage of books published throughout that

time, reviewing, for example, new novels, poetry, plays, and translations, also critical, scientific, historical, philosophical and political works, and books of travel and natural history. The reviewers (sometimes referred to as 'contributors') in these many subject areas wrote their reviews anonymously according to the literary convention of the time, and, except for the clues provided in the marked file of *The Athenaeum*, the identity of many of the reviewers remains obscure.

City University Library holds the unique 'marked file', a nearly complete run of *The Athenaeum*, annotated in the margins, with handwritten surnames of the reviewers against the reviews for which they were responsible. It is thought this was for the purpose of calculating payment.

The file was first brought to the attention of scholars through Marchand's seminal work on *The Athenaeum*.<sup>3</sup> Its importance as a source is also evident in all of the contributions in a recent special issue dedicated to *The Athenaeum* in the *Victorian Periodical Review* published by the Research Society for Victorian Periodicals.<sup>4</sup>

## *The project team*

Such a project requires extensive interdisciplinary collaboration. The preparatory work of extracting the information from the marked file has been carried out by our Belgium colleagues. The staff at City have been responsible for the design of the database, data entry and project coordination. To date proofreading and validation have been shared within the team. A consultative panel of subject experts, chaired by Diana Dixon from Loughborough University, has also been formed. Although its members have not as yet been convened as a group, it is anticipated that their participation will be valuable in future stages.

## *The system hardware and software*

The system has been built on a Dell 386 micro-computer. Since the ultimate size of the database could only be estimated, it was essential to have sufficient disc space or at least hardware that would be easily expand-

able. The equipment selected has proved to be fast and reliable.

The software is an entity type relational database, 'Tinman', produced by IME. Similar to hypertext software, it allows for links to be created between data elements of the database. Searching can thus be carried out in different ways. Searches can be selected from a menu which provides access to browse different alphabetical indexes (i.e., lists of authors, titles, reviewers, etc). Once in the index, the selection of an item and display of its individual record can serve to 'navigate' to other data sets. For example, from a selected title record giving the reviewer of the published work it is possible to display all of the other titles which that particular reviewer has reviewed. In addition to the 'browse and navigate' approach, a query can also be formulated by providing the searcher with a template or 'query by form'. This form-filling method allows for the use of Boolean operators and keyword searching as with any other text retrieval or bibliographic system.

To incorporate all the required links for searching, the design of the database needed to anticipate the different types of queries which were likely to be generated by the users. A fuller description of the database design and different types of searches possible is available elsewhere.<sup>5</sup>

### *The card index*

The first phase of the project, undertaken by Professor Wellens, consisted of compiling a card index from the marked file, noting along with the bibliographic details of every review, the reviewer's name as provided by the marked file. This was a considerable task of decipherment, particularly where the names were scrawled or cut when margins were trimmed in the binding process.

There are some 150,000 cards in strict date order. Each card represents a review of one book—where a review is of several works, a card has been made out for each individual book. Each record carries as much of the following information as *The Athenaeum* review provides: book title, (volume or edition where provided), translated or original title of book, author/editor/translator or designation as 'anonymous', reviewer's name as it appears in the marked file.

### *Data selection criteria*

In order to restrict the vast scope of the index, research has concentrated on enhancing the material with information on contributors/reviewers rather than on book titles and authors. So the following information is not included in the cards: publishing details, number of volumes of one work where the review addresses all, subtitles (all the words after the 'or', or colon, in a prolix extended title.) In the last case exceptions have been made where the full title was enlightening or potentially

provided a clue to authorship. (Some discussion of this follows.) The index is of only book reviews and reviewers and does not include the science, drama, music and art columns (also frequently 'marked') giving notices of meetings, forums, performances and exhibitions. Both long and short reviews in all these subject areas are, however, included in the index. There is often a crossover between the book reviewing and column journalism. For instance, F. G. Stephens, who reviewed books on art in the later part of the century, also reviewed exhibitions, and Henry Fothergill Chorley, the prolific book reviewer in so many subjects, also had a specialism in music and commented on opera. It might be possible to extend the database into these areas in the future. That the index exists as a 'live' database rather than a single, discrete print version might help make this a possibility. Figures 1 and 2 serve to illustrate the information included for data entry. Figure 1 is a specimen index card, not 'problematic' and restricted to essential information. Figure 2 consists of the same record as it appears on the data entry template of the database.

### *The database*

The current phase of the project, which has been running for two years, consists of entering the data from the card index onto the computer, while further research into the full or likely identities of the reviewers continues concurrently. Out of this research, a burgeoning 'authority file' is being built up of the contributors' names, dates and brief biographical details on each, and these 'authoritative' forms of the names from the marked file are incorporated into the individual records as the data for these is being input. Data entry has reached 1862. There is also a huge and continuous task of proofreading and editing.

The main difference between the card and the computer record is the inclusion of the 'preferred form' of the name. In our example, 'Frederick George Stephens' is entered with reference to the dates and biographical details contained in the 'authority file'. A relevant section of the authority file can be called up as a 'pop-up window' in the course of entering the data, to assist the inputter to ascertain which Stephens this is (if there is more than one contributor by that name) or, if there is only one, to check that the review was written within possible dates, and to enter the full name. In this case the appropriate entry in the authority file is shown in Figure 3. Figure 4 reveals the contributor's record details. The same biographical details used in compiling the index to complete the record will eventually be available to those searching the database.

In the course of transferring data from the card index onto the computer, and proofreading and editing the printouts of the growing database, a fuller 'picture' of the material has emerged. The computer generates its own indexes of the different data: i.e., authors, reviewers,

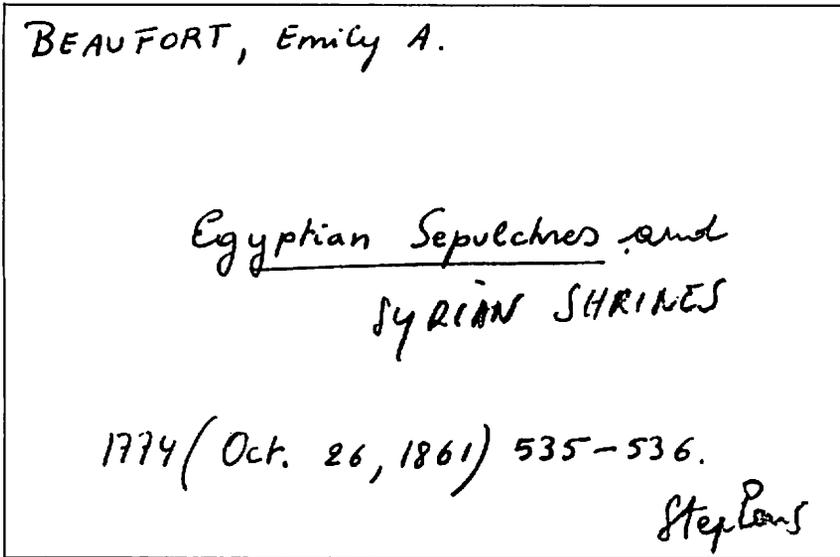


Figure 1. A specimen index card.

dates, which can be searched in relation to one another. It can show, for example, (at present only up to 1861), all titles written by a particular author such as Charles Dickens (there are 17), all titles reviewed by a specific contributor (2,407 reviews attributable to H. F. Chorley), or it can find all titles by a specific author (say Dickens) reviewed by a specific contributor (Henry

Fothergill Chorley) in a specific year (1861)—there is one (*Great Expectations*). These lists aid the research into the identity of the contributors, revealing patterns, for example, of subject matter tackled by a specific reviewer, or a cluster of dates.

Inevitably, in gaining the fuller picture, inconsistencies have been revealed and other problems arisen.

Title (and Volume) :Egyptian Sepulchres and Syrian Shrines

Author :BEAUFORT, Emily A.

"Please use validation facility to fill in the next field (press F10)"

Contributor (Preferred form):STEPHENS, Frederick George  
Contributor as marked :STEPHENS

To create another issue entry press <INS> on the "Reference field"

Reference :1774 (Oct. 26, 1861), 535 - 536  
Issue No. :1774  
Year :1861

"Do not fill in any more fields Keywords will be generated on storage"

Keyword :Egyptian  
Keyword :Sepulchres  
Keyword :Shrines  
Keyword :Syrian  
Catalogued :90/12/18

Figure 2. Data entry template.

Title (and Volume) :	Egyptian Sepu	STEPHENS, Frederick George
Original title :		STERLING, John
Translated title :		STEVENSON, Rev. Joseph?
Author :	BEAUFORT, Emi	STEWART, Sir Thomas Grainger
If the author is anonymous put "A"		STIGANT, William
Anonymous authors :		ST. JOHN, Horace Stebbing Roscoe
Author of :		ST. JOHN, James Augustus
Editor :		ST. JOHN, Percy Bayle
Translator :		ST. JOHN, Sir Oliver Beauchamps Coventry
"Please use validation facility t		STOTT*
Contributor (Preferred form):	STEP	STRICKLAND, Agnes
Contributor as marked :	STEP	STUART-GLENNIE, John??
		STUART, J.
		SWAN*
		SWEET, Henry
		SYLVESTER, James Joseph
		SYMONS*

Figure 3. Date entry authority file display.

### Problems in indexing authors/titles

There are sometimes inconsistencies in the way author's names are input, depending strictly on how these appear in *The Athenaeum*. For example, some Dickens titles (the 17 already mentioned) occur under Charles Dickens, but this is not the complete list. One more (*Barnaby Rudge*) occurs under C. Dickens and yet another title (*Master Humphrey's Clock*, Volume I) under Dickens' pseudonym, 'Boz'. The minor problem in this case can be solved by searching the database restricting the information provided/requested to the name

'Dickens' on its own. The entry under 'Boz' can still be found in this way because the database has a special facility which can link pseudonyms with real names. The same system can be used for searching book titles and contributors/reviewers.

Since that link between 'Boz' and Dickens is not made in *The Athenaeum* itself, the project researchers have to establish such links. One of the editorial questions consequently is: to what extent is it feasible, given time constraints, to provide information about authors gained from secondary sources? (The main research has

Contributor (Preferred):STEPHENS, Frederick George

Dates :1828 - 1907

Characteristics :Art critic. Member of the Pre-Raphaelite 'Brotherhood'

Source :DNB

Figure 4. Contributor's biographical details.

been into the identity of contributors.) So far, this is done where the work is well-known, where it would be an obvious mistake not to do so. So a link must be made between Boz and Dickens, between Ellis, Currer and Acton Bell and the Bronte sisters, and, in the case of 'anonymous' authors who are now known, Mrs Oliphant must be substituted for 'anon' as author of 'The House on the Moor' and Mrs Gaskell as author of 'The Moorland Cottage'. It does, however, seem significant to record that for the purposes of original *Athenaeum* readers, these books were anonymous. How widely must this procedure be applied to make less obvious links? Many more works which appear to be anonymous judging by the *Athenaeum* review alone, have, or might have authors who are now known, and our data entries must remain provisional.<sup>6</sup>

Similarly, there are many pseudonyms not linked to any other name. Many pseudonyms masquerading as 'feasible' surnames must have escaped us. Simply scanning the author indexes is astonishing in the extravagance and range of 'legitimate' names! Yet some of these can safely be assumed to be pseudonyms ('Sea Story' by 'Land Lubber'). These are entered into the author 'field', often in quotation marks, and dispersed throughout the indexes in alphabetical order. It is also intended to make an additional file putting these together as a category of 'anon'.

Where an author, especially of novels, is anonymous, *The Athenaeum* often provides other titles by the same hand, in an 'author of' review heading. These are entered into the computer in a special field which links them all together, and providing one has one title, one has access to the rest. Moreover, *The Athenaeum* sometimes offers a title, the author of which is named, as a sequel to an earlier work. This information can be important. The earlier work, which if ever it was reviewed in *The Athenaeum*, can be searched by title on the computer, often proves to be anonymous and sometimes linked with a further series of anonymous works by the same author.

So 'Earnestness' by Charles B. Tayler was reviewed in *The Athenaeum* in June 1850. It is cited as the sequel to 'Thankfulness'. 'Thankfulness', reviewed in April 1848, had an anonymous author, also responsible for 'Records of a Good Man's Life'. All these records can thus be linked under one author.

These special anonymous cases raise questions for the indexing policy in general. The original policy, in generating the card index to *The Athenaeum*, was to restrict the scope of the index in certain ways already outlined, in order to keep it manageable. For these reasons, a short title format was preferred. However, to have pursued this policy rigidly for the sake of consistency alone, would have meant omitting significant

details and thus failed to make certain links between titles and to establish authorship. Some information is therefore provided in such cases.

Sometimes information from an extended title proves useful in identifying the nature of a publication which might remain opaque or misleading given only a short title. For example, 'Seasons with the Sea-Horses' by James Lamont in *The Athenaeum*, 2 March, 1861, is not a poetry anthology as one might imagine; the full title is 'Seasons with the Sea-Horses; or, Sporting Adventures in the Northern Seas' and is a gruesome account of the pointless slaughter of some 600 walrus, 70 polar bears, etc., in the Arctic, by a defeated Member of Parliament in the late 1850s. There is more access to this unfortunate document if the title is catalogued in full. If one does a keyword search on the computer for 'sport' the system finds 'Sea-Horses', also many other references to hunting and one 19th-century attitude towards wildlife.

The normally short title format may also be enhanced by additional material on secondary sources. For example, *The Letters and Works of Lady Mary Wortley Montagu*, edited by Lord Wharnclyffe, Third Edition, Volume II, is reviewed in *The Athenaeum* in October 1861. This volume, *The Athenaeum* tells us, carries a 'New Memoir' by W. Moy Thomas. William Moy Thomas is also a frequent contributor/reviewer in *The Athenaeum* at this time, therefore it may be useful to researchers to cite this example of his other published work.

### *Problems in indexing contributors/reviewers*

By far the largest difficulty in indexing *The Athenaeum* has been identifying the contributors/reviewers, anonymous except for their 'marked' surnames, and providing accurate information in respect of 150,000 records. It has necessitated continuous liaison between the researchers in Belgium and staff at City responsible for data entry and proof correction. Much has had to remain provisional. Editorial conventions have had to be agreed for expressing different degrees and types of doubt. As patterns of contribution have been seen to emerge, many records left provisional are then 'filled in', using the links made by the database.

The 'authority file' to which the data inputter has access contains an alphabetical list of all reviewers' names so far established in their 'preferred' forms.

Often the marked form of a name as it is accurately recorded on the index card does not correspond with the 'preferred' form. For example, is Lawrence, Lawrance, or Laurance the same person as Hannah Laurence, reviewer of books on local history, dates 1795-1875 (Source: *Who Was Who/Dictionary of National Biography*)? If, as was decided, she was, then the variant spellings must be incorporated into the authority file and also offered on each individual record. Or is Beecher or

Bucher the same person as Alexander Bridport Becher (1796–1876) (Source: *National Union Catalog*)?

There may also be reviewers on the authority file with the same surname, contributing reviews to *The Athenaeum* at the same time. One example is the ubiquitous Henry Fothergill Chorley, dates 1808–1872, (Source: *DNB*), the most prolific reviewer for *The Athenaeum* ever, yet writing at the same time as his brother John Rutter Chorley, dates 1807–1867, (Source: *DNB*). Writing style and choice of subject differentiate them, and the occasional provision in *The Athenaeum* of initials for J. R. Chorley make it possible to be sure for certain reviews, but for many other reviews, data entries must have a question mark. All such entries must eventually be extracted and go to a scholar who specializes in the Chorleys. (The consultative panel of subject experts will handle such questions.)

Other well-known reviewers with the same name are Lady Sidney Morgan, dates 1776–1859, (Source: *WWW*) and her husband Sir Thomas Charles Morgan, dates 1783–1843, (Source: *Webster's biographical dictionary*). Again, it is usually possible to tell them apart on subject matter, and due to their relative eminence they prove amenable to research. Still, all such conjectural entries must be re-checked. Other pairs of reviewers are more problematic. David Masson, 1822–1907, (Source: *WWW*) and George Joseph Gustave Masson, 1819–1888, (Source: *DNB*) are difficult and sometimes impossible to tell apart since they wrote on many of the same topics. In some records, one of the names is entered speculatively, in others both names are entered as equally possible. There are many 'pairs' of reviewers like this. And, there are six Browns, five Bells, four Cunninghams, four St. Johns, four Watts, three Thomases, and three Whites, many of whom could be confused!

However, for the later 19th century, the problems of identifying contributors lessen. For the first 40 years of *The Athenaeum*, many reviewers wrote prolifically on miscellaneous subjects. As the century went on reviewers became more specialized, making identification easier. Yet late in the century the number of books reviewed per issue greatly increased. At the *Athenaeum* project we have so far entered data up to 1862. As the pattern of reviews changes to generate a greater number of records per issue, then per volume, this may create different problems.

### Future prospects

It is estimated that the data entry will be completed in eighteen months to two years' time. In the meantime we will be concerned with evaluating the system with users to ascertain:

1. in what form the index should be made available;
2. what impact such a tool will have on scholars.

Unlike previous printed indexes, the online environment allows for the manipulation of data in a variety of ways. Although elements of the database could be published in a traditional form, the ideal would be to make the full system accessible to researchers.

Because of the scope of the subject coverage, the database is potentially of interest to a wide range of scholars. Undoubtedly it could enrich existing areas of research as well as promote new ones. *The Athenaeum* had a wide circulation both in the UK and abroad. The index would not only be useful to libraries who subscribed to the journal, but could equally be used as an independent resource.

The project team is now endeavouring to attract the necessary financial support for the next phase to make this unique tool widely available.

### References

1. Beare, Geraldine. Indexing the *Strand Magazine*. *The Indexer* 14 (1) Apr. 1984, 8–13.
2. Houghton, W. E. (Editor), *The Wellesley indexes to Victorian periodicals*. Vols. I–IV. Toronto: University of Toronto Press. 1966–88.
3. Marchand, L. *The Athenaeum: a mirror of Victorian culture*. New York: Octagon Books. 1971.
4. Athenaeum Issue 1. *Victorian Periodicals Review*. 23 (1), 1990.
5. Hancock-Beaulieu, Micheline. *The Athenaeum: the creation of a computerized index and database to a Victorian literary periodical for researchers in the humanities*. In: *Information technology and the research process—proceedings of a conference held at Cranfield Institute of Technology, UK, 18–21 July 1989*. Edited by Mary Feeney and Karen Merry. London: Bowker-Saur, 1990. 278–87.
6. Hiller, M. R. The identification of authors: the great Victorian enigma. In: *Victorian periodicals: a guide to research*. Edited by J. Don Vann and R. T. Van Arsdell. New York: Modern Language Association of America, 1978. 123–48.

---

*Micheline Hancock-Beaulieu is director of The Athenaeum Indexing Project in the Centre for Interactive Systems Research, Department of Information Science, School of Informatics, City University, Northampton Square, London EC1V 0HB.*

*Susan Holland is responsible for the database and editorial matters for the project.*

---

### Freelance:

—one who gets paid per word, per piece, or perhaps. Thus defined by Robert Benchley, as quoted in *Selected letters of James Thurber* (Hamish Hamilton, 1982).