Indexing in the context of microform publishing

C. E. Chadwyck-Healey

Director of Chadwyck-Healey Ltd, Publishers

I would like to say how much I appreciate being invited to address this Society, because indexes play a central role in microform publishing.

I should warn you that I use the term index in its loosest sense, interchanging it with catalogue and even handlist or guide, since what I am really talking about are the searching tools needed to make many microform publications usable.

Microform and microfiche

Incidentally microform is a generic term that covers both microfilm—that is rolls of 35mm or 16mm film—and microfiche—6in. by 4in. sheets of film.

The problem is that people are frightened of microforms—which is understandable. Microfilm in particular is what I call a closed format. The film is wound on a reel; it is impossible to glance at what is on it. Manually unwinding the first 10 ft. to see even the first frame which may have an eye-legible title on it, is laborious and time consuming; yet that one reel could have up to 2,000 pages on it. Maybe there is an index at the end of the reel—perhaps part of the book or periodical that has been filmed, but it takes an appreciable amount of time to wind through to the end of the film to find the index, and then you may find only that the index refers you to an issue or page on another reel. You then have to rewind the reel, load on the other one and begin again—an exercise in total frustration.

Microfiche is an open format—each 98-page unit is labelled with an informative title strip giving title and page numbers contained on that microfiche. Individual frames may have eye-legible captions so that you can get a good idea of the content of the microfiche without even putting it in the reader. Furthermore, the reader itself is much easier to load and unload than that for microfilm.

You will already have realized that searching for material on microfilm is a linear process—you can’t jump from page one to page 600, you have to wind through every page to get from one to 600, or vice versa. Microfiche is random access, much more akin to a book where you can turn from one page to another without having to work through the intermediate pages. First, you can select your fiche at random and then once the fiche is in the reader you can move the platen directly from one frame to another in any order required.

These physical differences are fundamental to the acceptability or otherwise of microform as a medium for reproducing information. I should add that the reproduction of visual material on microfiche has its own special characteristics which I will later consider.

I think my description makes microfiche sound a much more accessible format than film, and this seems to be borne out by the rapidly growing popularity of the microfiche format in comparison with microfilm. We now have thousands of libraries with their catalogues on microfiche, and in your local garage you are likely to find that parts for your car are listed on a computer output microfiche supplied to the garage by the manufacturer. It is fiche, not film in cassettes, because fiche are easier to use and the readers are less expensive.

We now produce far more microfiche than microfilm but our first two publications were on microfilm, one of them producing some interesting indexing problems.

Archives of British publishers

This was the first series of archives of British publishers. I felt that there would be a keen interest in the records of the important publishers—if these records existed, could be found, and be made available to us. After having contacted every publisher who was active in the 19th century I began to realize that there were more records in existence than might have been assumed, though the number destroyed by fire is distressingly and surprisingly large. However, it was largely due to the generous attitude of Allen & Unwin and Cambridge University Press that the project got under way, because we were quickly given permission to microfilm the letters of George Allen and Swan Sonnenschein, and the archives of Cambridge University Press in the University Archives and in the offices of the Press itself.

Allen and Swan Sonnenschein’s records were in the offices of Allen & Unwin at the very gates of the British Museum. Were they important? Yes, primarily because Allen was John Ruskin’s engraver and was set up by Ruskin as a publisher to combat the iniquities, as he saw it, of conventional commercial publishing; and Swan Sonnenschein was the publisher of the first English editions of Marx and Engels—which is why a set of the microfilm now resides in the Marx-Lenin Institute in Moscow.

*Text of the address given to The Society on 12 July 1980.

The Indexer Vol. 12 No. 2 October 1980 73
Each archive contains about 50,000 out-letters from the publisher. Tragically, no incoming correspondence survives. Unwin's solicitors destroyed all Marx's letters when Unwin took over Swan Sonnenschein and Allen in the 1920s—a remarkable act of vandalism that lawyers are only too prone to.

Each volume contains about 1,000 letters—an index was clearly needed, but what was the best way of making it?

We found that each volume had its own index, compiled presumably by a clerk in the office on completion of the volume. We started to compile an index of people and firms to whom the letters had been written, using of course the microfilm so that the fragile volumes could be handled as little as possible.

After a few months of head-burying we faced up to what we should have realized immediately—that the original indexes were almost useless in that the clerks had misread the names in so many cases or had carelessly referred to the wrong pages. There was no way round it; we had to chuck out all our work and start again. I have regarded original indexes with great suspicion ever since.

The logistics facing us were formidable; 50,000 references in each of these two archives out of the six we had microfilmed and were already receiving orders for. We had sheets of perforated card made which could be torn into small 2in. by 1in. cards, and special aluminium trays 4ft. long made to take these cards. We then supplied typists working in their own homes with microfilm readers so that they could type names and page references straight off the microfilm on to the sheets of cards. Our editor then sorted the cards, picked up anomalies in spelling, initials and so on until eventually we had the cards ready for typing for the index which is typewriter-set.

And of course the indexes are worth their weight in gold. Many libraries buy only the indexes. A scholar can use the index to see if there are any references to the person he is interested in. If there are he can borrow the microfilm from another library or even ask the other library to make photocopies of the relevant pages.

This does not help us, since the price of the indexes does not really reflect the cost of producing them, but we know that there are only a limited number of libraries which can justify buying the microfilm and we want to see the archives used as widely as possible.

Now, after eight years, indexing of the archives of publishers continues. We have had Alison Ingram in Cambridge working continuously since 1975. She has compiled indexes to the archives of Richard Bentley, Longman, Grant Richards, and now Harper and Brothers—the first of a series of American publishers' archives. All the indexing is done from the microfilm, which I think says something for the quality of reproduction obtainable with microfilm. When she indexed part of the archives of Richard Bentley at the University of Illinois she found that she was able to identify on the microfilm names which the library staff had been unable to identify on the originals.

Another point about the Bentley Archives is that Richard Bentley the third had sold the firm to Macmillan in 1898 to pursue his own antiquarian interests—amongst other things he published a brief history of Slough with a good deal of information about its rainfall. In the 1930s his widow disposed of the archives piecemeal; some went to the British Museum, some to the University of Illinois and a third part to the University of California at Los Angeles (which also has the superb collection of first editions of Bentley novels in mint bindings sold to Michael Sadleir by Richard Bentley III).

Using the archives scattered in this way was almost impossible. Royal Gettman, who wrote a history of Bentley, told me that whichever part he was using he always needed to look at items in the other parts. The microfilm is in three separate sequences and it is only through the index that we have been able to consolidate the entire archives into the original form in which they should have remained.

**The Diary of Beatrice Webb**

A fascinating and important archive for which we compiled an index was the Diary of Beatrice Webb. We published the original holograph and a typed transcript on microfiche with the London School of Economics. They undertook the compilation of the index though we were responsible for its printing. In fact we became involved in the compilation as well and certainly share responsibility for it, and for any shortcomings, of which there were a few—for example, Eleanor Marx was described as 'the wife of Karl'. This was a point picked up by a reviewer and one has to accept such justified criticism. If there were no indexes, or better still if every index was perfect, what would most reviewers do? They know that if they don't have time to read the book they can always criticize the index—or comment on its absence!

**BBC Radio news index**

Our most ambitious indexing project to date ended in grief. We have had a close association with the BBC for five years; we have microfilmed the card catalogue of authors and titles in the radio and television drama libraries—probably our best-selling microform title—and have microfilmed the texts of the Radio Nine O'Clock News for every day of the War—1939 to 1945. The success of this led us to the publication of the contemporary six o'clock radio news on microfiche issued every quarter. Clearly this had to be indexed if it was to become an everyday reference tool and equally

74 The Indexer Vol. 12 No. 2 October 1980
clearly the index was going to be quite a demanding exercise—100 pages a quarter of double-column A4, containing in one year 100,000 references.

In fact it went better than we might expect in view of our inexperience of a publication of this type, but the lack of a thesaurus and the inexperience of our indexers did create some extraordinary inconsistencies and anomalies—especially in the first couple of issues. Unfortunately it was these issues that were appraised by reviewers, but they were kinder than perhaps we deserved. An example quoted by Peter Thomas, Librarian of the Consumers' Association, in a review in the Reprographic Quarterly, mentions the Good Food Guide being indexed under Restaurants and the Michelin Guide under Food. Prince Charles appeared under Wales, Prince of, but unfortunately there was no cross reference from Charles.

Our standards of indexing improved very quickly. Unfortunately our sales didn't. The enthusiasm of ourselves, the BBC and even Peter Thomas, for the project was simply not matched by libraries here and overseas.

The microfiche and the index started in 1978 and ended at the end of 1979. I wonder what librarians in the future will make of this short-lived ray of light briefly illuminating a sea of oral darkness.

New light on feudal England

We have just announced a new history publication which breaks entirely new ground in providing access through computerized indexes to large collections of historical documents.

A group of researchers at the Department of Social Anthropology here in Cambridge have fed into the University computer 7,000 pages of historical documents of the Essex Parish of Earls Colne from 1400 to 1750. Earls Colne is one of the best-documented villages in Britain, and through these documents it is possible to reconstruct English economic and social history from the middle ages to the eve of the industrial revolution.

The entire texts of the documents in their difficult-to-decipher hands have been transcribed into the computer and coded for access. We are publishing the texts of the documents which are of course now easily readable in modern type, and various indexes are being published so that it is possible to trace one person or a family, or a house or piece of land, through a long period, as references appear in a diverse range of documents. Scholars are encouraged to ask for special searches to be carried out by the team before they disband, but the published indexes are sufficiently sophisticated for most needs and allow an extraordinary degree of accessibility to such a huge and intractable mass of documentation.

Already historians associated with the project state that the evidence appears to show that feudal England of the late middle ages was quite different from what has been assumed. Instead of simple peasants tilling the same piece of soil for the whole of their lives, hardly ever venturing from their village, there emerges a picture of a highly mobile, almost unstable, society with people moving from one village to another throughout England. Land and houses were being bought and sold frequently; endless marital disputes, drunkenness and petty crime. It sounds so much like modern society that perhaps it is the key to why the industrial revolution started in England. This is perhaps an outstanding example of important historical research that has been made possible by sophisticated indexing.

Art exhibition catalogues

Much of our activity over the last five years has been in the field of fine art, decorative art and architecture.

Our main publication is an open-ended series of art exhibition catalogues reproduced on microfiche. We now have over 4,000 separately available titles, ranging from French Salon catalogues of the 18th and 19th centuries to 1980 catalogues of the Guggenheim in New York. Each catalogue is identified by a three-part number, which proved invaluable when we came to computerize all our sales records for this project. In our sales brochures the titles are listed by keyword with cross reference where necessary. An artist's name in a title is automatically chosen as the keyword with cross references if there is more than one name.

We also publish annual lists of titles with a subject index and an index of agents of publication, i.e. museums and galleries. Fortunately for us and the users of this series a more substantial index also exists. It is the catalogue of the Art Exhibition Catalog Collection in the Art Library of the University of California, Santa Barbara. This library has been actively collecting art exhibition catalogues for a number of years and now has a collection of over 30,000 titles. Bill Treese, the art librarian, decided to design a cataloguing system with information stored in a computerized database with print-outs available to other libraries in the University of California system. Very quickly the print-outs became unmanageably large and the demand for copies also grew. We published the catalogue on microfiche and added an update to the base catalogue every six months. It now consists of the base catalogue, containing about 20,000 catalogues, and five updates containing some 10,000 catalogues. We will shortly reissue the whole catalogue in a consolidated form. The computer tapes are sent to us from California and the computer output master and copies are made in the U.K. Ninety per cent are sold back to institutions in the U.S.A.; I think it's a nice example of coals to Newcastle. However, the catalogue has an added importance for us: the library takes all our exhibition
catalogues on microfiche and indexes them in the catalogue. They are identified as being on microfiche and our three-part serial numbers are also given, with a prefix SH—standing for Somerset House, our company in the U.S.A. This means that our microfiche titles are being indexed by experienced art indexers using a sophisticated and carefully worked out system, and this adds greatly to the usefulness of the microfiche editions.

The Catalog is divided into two main sections, the Agency Index and the Subject List. The Agency is the publisher and the titles are listed by publishing agents. Also given are the dates of the exhibition, the library’s accession number and some bibliographical information. In the Subject List each title is entered under one or more of the following five main descriptions: Artist, Style or movement, Medium, Subject, Collection. Each main entry can be sub-divided five times into: Country, City, Date by century, and up to two Specific Dates, e.g. 1840 to 1860. It is essential that the main description be qualified in this way. Thus, a subject such as Portrait Painting, far too general to be of much use on its own, can be combined with subsidiary descriptors; Country, Century, Dates; to produce, Nineteenth Century French Portrait Painting 1880-1890.

One of the advantages of using a computer is that the information can be accessed in different ways. Thus, we could produce an index of authors, or cities where exhibitions have taken place, or a subject index listing the main and subsidiary descriptors. Another advantage is that the information can be made available on-line through the terminals which most large libraries now have.

Scanning prints

Sometimes a microform publication is made possible by the existence of an index or catalogue. An example is a microfilm publication embracing both history and art: the English Cartoons and Satirical Prints in the Department of Prints and Drawings in the British Museum. We have reproduced on microfilm the 17,000 prints listed in the monumental 12-volume catalogue (totalling 13,000 pages) compiled by Stephens and George. The first volumes were published in the late 19th century; then there was a gap until Dorothy George took up the work in the 1930s, completing the catalogue in the 1950s. Each print is given a lengthy description, with good indexes in the later volumes, and the prints on microfilm arranged in the same chronological order as in the catalogue. In this case microfilm has the advantage of a large frame-size suitable for reproducing the larger prints. It is also fascinating to be able to wind through the prints, like a huge comic strip of English history. In this case the ability to scan hundreds of images in a short time is made possible by the microfilm format. To see the same number of prints in the British Museum itself would take 20 times as long. But still there would be no point in microfilming a collection of prints that had not already been identified, described and indexed. So here the catalogue was essential.

We realized that microfiche in particular is an ideal format for reproducing large collections of visual material. First, it is an inexpensive method, cheaper than slides or books; second, the quality of reproduction is high, mainly because it is a straightforward photographic process with none of the complexities of half-tone printing; and third—and this is most surprising—the images are more accessible on microfiche than in any other format. For example, I can hold up and scan a colour microfiche with 60 images in a few
seconds and see enough to know whether it contains material of interest to me. I can put it in the reader and examine each image blown up to full size. I can then go a step further and project the images for a class or seminar. But once again the advantages of the format are nullified if there is no index to the images on the microfiche.

Collections of visual material

We are microfilming four major picture collections, all in the U.S.A. The Index of American Design in the National Gallery of Art, Washington, contains 15,000 paintings of American decorative arts and crafts from settlement to 1900. To make publication possible we had to arrange the entire collection into recognized subject-categories, and catalogue and index each item. This proved unexpectedly expensive and time consuming; the 700-page cloth-bound catalogue is the most expensive book we have ever produced.

We are microfilming a collection of 87,000 photographs in the Library of Congress—the Farm Administration and Office of War Information photographs showing all aspects of life in America from 1935 to 1946. In this case the Library of Congress has arranged the captioned photographs in six regions under 1,300 subject headings. We are publishing the subject headings in a printed guide and in this way this huge collection becomes accessible.

Another photograph collection made usable by the existence of a detailed catalogue is the Vandamm collection in the New York Public Library, Library of the Performing Arts. This is a collection of 26,000 photographs of stage productions in New York from 1919 to 1961. We are also reproducing on microfiche the card catalogue of playwrights, arrangers, producers, actors, set designers, lyricists, choreographers and costumiers. One can find out from this catalogue in which productions a person took part and then go straight to the photographs of that production on microfiche. In this case the catalogue itself is valuable even without the photographs, since it is a unique reference work for the history of the New York stage in recent times.

Parliamentary Papers

Two major indexing and cataloguing projects will take up much of our time and efforts over the next few years.

We are publishing a new microfiche edition of the House of Commons Parliamentary Papers for the 19th century. It is four-and-a-quarter million pages on 46,500 microfiche. There are 22 libraries in the U.K. supporting this project, and we are selling copies overseas. We have to keep detailed records of the location on microfiche of the 80,000 separate documents in the collection. It seemed that we should go further and produce new subject indexes for the entire century. We are now trying to draw up suitable subject headings. Eventually a computer programme will be written into which all 80,000 titles will be fed, with their subject categories. It has been suggested that we make the database available on-line—maybe it will be the first great 19th-century history collection to be accessible on-line.

At present we are faced with more problems than solutions. So much happened in that century that one questions the wisdom of collecting it all into one index; old subjects fade away, new ones appear as the century progresses. Terminology changes subtly. Can a term used in 1805 be used in quite the same way in 1890?

We are gaining insight into the methods of 19th-century indexers since we are interested in how they dealt with these documents even though their methods require some getting used to; this is perhaps a good example of microfiche publication fostering ancillary indexing.

Government publications

Lastly, a completely new venture for us, our forthcoming Catalogue of British Official Publications. Not Published by H.M.S.O. This is a new catalogue which will cover the many thousands of publications produced by Government departments, research institutes, nationalized industries and other official bodies, which are not published by H.M.S.O.


The catalogue, designed by ex-Aslib information specialists Ruth Finer and Judith Collins, is a listing of titles by departments with a comprehensive index of authors and subjects. We are also able to use the new House of Commons Library thesaurus for the index. I believe that this catalogue will be one of the most important new reference works published in recent years. Furthermore the documents will be available from us on microfiche so we will in effect be a central clearing house for the publications of hundreds of publishing agencies large and small.

Customers will be able to place standing orders for documents on specific subjects, so each document is given two sets of subject references, one for the index, the other for the standing order subject categories. This information will be stored in a computerized database. This catalogue is another indication of the close relationship between microform publishing and the production of indexes and catalogues, and our growing dependence on computers is in contrast to our first efforts with cards and aluminium trays.
I foresee a time when we may spend more of our resources on indexing than we do on microfilming. This is why I have been so glad to have had the opportunity to make our publishing interests and activities better known to you.

Discussion

Elizabeth Wallis asked whether there was a change in the type of qualities required of indexers working with these new techniques. Mr Chadwyck-Healey said different qualities were preferred for the different types of index; for the publishers' archive index, a knowledge of English literature and publishing rather than high-powered indexing technique which would be frustrated; whereas the catalogues indexed were not highly technical but needed qualified library cataloguers familiar with computer techniques. The key person is the in-puter of information; if their work is inaccurate, the indexer's work comes to nothing. The index to the 19th century Parliamentary papers was being done by a specialist in 19th century Parliament.

Doreen Blake asked about the life expectancy of books and microfiche or film. Humidity and room temperature are important for storage. Microfiche is susceptible to scratching; a new form, Diazol, has the image inside the film and therefore safe from scratching, but is impermanent as it is affected by light over the years. Books may last less than 100 years because of the acid in the paper; microfiche should last longer. For runs of more than 1,000, conventional printing methods are cheaper. Microfiche may also be affected by fungus in the tropics, and must be stored in drawers sprayed with chemicals to get rid of the fungus. Mr Chadwyck-Healey did not consider that the book would be replaced by videodiscs or microfilm, but that all three would survive together. A European firm had been publishing monographs on fiche for 25 years, but the book had not been superseded.

Peter Greig asked about collections of microfiches sold to libraries. Mr Chadwyck-Healey considered these rather bogus projects, good for marketing only. A thousand disparate titles could be filmed and sold as a collection to libraries throughout the world. Some such collections were very strange amalgamations of material. Mr Greig considered this a cynical view, and saw value in some Canadian projects whereby specialist bibliographies were sold to libraries as collections. It was true though that libraries might already have some of the titles in stock in original form, and would be duplicating these if buying the collection.

Mr Greig asked whether the indexes published by Chadwyck-Healey were in microform. Most are preferred in conventional book form.

Elizabeth Wallis pointed out that in terms of microfiche, sales of 500 indicated a best-seller. Marketing was safe however, as a microfiche project is announced, orders taken, and reprinting only then undertaken; it is not speculative. Joyce Skinner asked whether it will be possible to buy single units at normal book prices; it will be, at about £1.50 per individual catalogue. Main sales, however, remain to large libraries placing blanket orders.

The near future

Soon the ability to program and use microcomputers will be as important as being able to read, write, type, drive, or use the telephone, and the likelihood is that before very long employees may communicate with their work rather than commute to it, according to two forecasts of the social and economic consequences of word processing reported in Information hotline 11 (11) Dec. 1979. The value of word processing lies in increased clerical-labour productivity; the value of communicating word processors is a bonus of time.

Motor cars and the network of roads needed for them to transport people and goods are the analogy offered for the dependence of microprocessors on a developed network infrastructure. At the moment, incompatibility between the signals from word processors of one manufacturer and those of another manufacturer is the great barrier to communicating with word processors. Attempts, however, are being made to produce conversion packages and so reduce babel among computers.

For access to the most recent coinings of technical terms use is increasingly being made of terminological data banks held on computers. The authors of 'The translator as a computer user', H. L. Somers and J. McNaught (Incorporated linguist 19 (2) spring 1980, 49-53, 44 refs) foresee personal computers also holding dictionaries and enormously reducing searching time for translators, and, we may add, for indexers also.

Online information

The fourth International Online Information Meeting is to be held at the Cunard Hotel, London, 9-11 December 1980. The programme of papers will be supplemented by an exhibition of information products and services and displays of current technology. Details from the Organizing Secretary, Online Review, Learned Information (Europe) Ltd, Besselsleigh Road, Abingdon, Oxford OX13 6EF.

Proceedings of the First, Second and Third International Online Information Meetings, 1977, 1978, 1979, are available from the same address for £15, £17.50 and £25 respectively.

The Indexer Vol. 12 No. 2 October 1980