

Newspapers on CD-ROM

Introduction

This paper concentrates on British newspapers available in CD-ROM form. It looks at the characteristics of this form of publication of newspapers in comparison with the alternatives of hard-copy, microfilm and online, and discusses the benefits and limitations of the medium from the point of view of libraries and end-users. It does not provide a detailed description or evaluation of the individual products, since this is available elsewhere (1), and since the products themselves are still evolving, nor does it cover overseas titles or indexes to and abstracts of newspapers available on CD-ROM.

British Newspaper Titles Available

At the time of writing five CD-ROMs of UK newspaper titles are publicly available, some containing the text of more than one newspaper. The titles and the years of publication covered, together with publisher/distributor details are:

The Times/The Sunday Times 1990-
Times Network Systems/Chadwyck-Healey

The Guardian 1990-
Chadwyck-Healey

The Independent/The Independent on Sunday 1989-
FT Profile/Chadwyck-Healey

The Financial Times 1990-
FT Profile/Chadwyck-Healey

The Northern Echo 1989-
The Northern Echo

With the exception of the Northern Echo the data from each year of a title is usually contained on a separate disk.

Geoff Smith

British Library Newspaper Library

Newspaper from the point-of-view of Users and Libraries

To producers and most readers of newspapers a newspaper is primarily something bought on the day of publication, providing, on a daily or weekly basis, reporting and comment on current events, backed up by features, listings and other 'recreational' material. For both the normal expectation is that the newspaper will be discarded after use.

To librarians and to researchers, however, newspapers have long been recognised as valuable tools and information sources many years after their initial publication. The British Museum and later the British Library have systematically collected newspapers for the last 150 years and runs of local newspapers are among the most highly valued parts of public library local history collections.

Initially the only form in which newspapers could be kept for long term use was as hard copy or originals, usually by binding the collected issues of a title into volumes a month or year at a time. After World War II microfilm copies of newspapers became recognised as an acceptable surrogate for the originals offering the possibility, through the controlled storage of master negatives, of archival quality film of long term preservation beyond that likely to be feasible (at least economically) of the originals.

The last decade has seen the growing availability of the text content of certain newspapers in electronic database form on online hosts such as FT PROFILE, bringing to those users able to afford it access to very large amounts of newspaper information

through the power of information retrieval software. The capability of CD-ROM to handle large full-text databases has enabled the publication of newspaper text information in this medium and in the last sixteen months has made electronic access to it available for a more general audience.

Given the availability of newspaper information in these four media, the choice for libraries and for users is dependent on a mixture of organisational or user needs and on economic and technical considerations.

Historically, national and public libraries have concentrated on collecting and preserving for use the original newspapers or microfilm copies, with subject access limited to the availability of published or in-house produced indexes, while commercial and other users, for whom the information content of stories is the primary concern, have opted either to build collections of newspaper cuttings or to use online databases where the data range and title coverage available meet their needs.

Newspapers on CD-ROM

Where then do newspaper on CD-ROM fit in the spectrum of choice and what benefits and disadvantages do they offer compared to the alternatives?

Costs

A comparison of the costs of annual subscription prices for CD-ROMs with those of microfilm or hard copy plus printed index shows the products to be competitively priced.

	CD-ROM	Microfilm	Hard copy	Index
<i>Times/Sunday Times</i>	£595	£700	£165	£450
<i>Guardian</i>	£495	£330	£145	£470
<i>Independent/Independent on Sunday</i>	£499	£610	£170	
<i>Financial Times</i>	£900	£585	£170	£570
<i>Northern Echo</i>	£100			

The prices shown are for 1992 subscriptions (some index prices based on conversions from dollar prices at the time our subscriptions were paid). The hard copy

prices are those of the cover-price of the individual issues and do not include the cost of binding or other form of keeping the copies fit for long term use.

In the case of *The Guardian* and *The Times* these prices have been cut significantly since their original launch, while for all of the titles discounts are available to public sector educational organisations which reduces the price further for those users.

I have not attempted to compare the costs of online use since these are variable, being based on connect time, display charges and telecommunications costs, with searching across multiple titles and years the norm.

Characteristics of CD-ROM in Relation to Original Newspapers

The first and obvious thing to say about CD-ROM versions of newspapers is that they are not a substitute medium for the primary purpose of newspapers identified above, i.e. the current provision of daily or weekly reporting of current events and the associated editorial, feature and leisure material that accompanies it. Nor, like online files, do they provide a full archival electronic equivalent of the published newspaper.

Instead they provide an electronic archive of the main information content of newspapers, the news stories, feature articles, editorial and other matter but excluding a great deal of the material that contributes to the identity of individual newspaper titles and which is of interest to both current readers and future researchers.

Exclusions

Material excluded from CD-ROMs varies from product to product, although some exclusions are common across all titles.

In most cases photographs and other image and graphic material are currently excluded. This is likely to change as the technology develops. The 1991 end of year

disk from *The Times* contains several hundred photographs but these represent only a small proportion of those published in the year and so far these are not well integrated with the articles which they illustrate. The announcement of the forthcoming CD-ROM version of *The Economist* says that it will include 'graphics, charts, maps and tables where they are thought to be important to the article'.

Advertisements are excluded in all cases, including the births, deaths/marriages announcements of major interest to future generations of family history researchers.

Most "recreational" material is excluded, including puzzles, crosswords, cartoons, TV and sports fixture listings, recipes, and weather reports. Many of these are features which beyond their immediate appeal to the purchaser of the original newspaper are of value and interest to future researchers.

In other cases individual titles have adopted differing exclusion policies, often based on differing interpretations of copyright issues. Thus readers' letters are excluded from *The Guardian* but included in the others. Where individual writers retain copyright their articles may be excluded. News agency stories are often excluded, while tabular material such as stock market prices is generally excluded, even from *The Financial Times*.

In general the exclusions are similar to those in online files, almost certainly because the online and CD-ROM files are created from a common source. There may also be an underlying assumption that the information needs of CD-ROM users will be the same as those of the commercial and similar users of on-line systems, that is primarily factual information from recent years and from the 'quality' newspapers. Experience at the Newspaper Library shows that the long term historical research use of newspapers is often different from, and much more varied than, this.

Benefits and Limitations of CD-ROM

Benefits

The benefits of CD-ROM as a medium for the publication of, and access to, newspaper material are similar to those of CD-ROM in general.

To the individual end-user they offer: the power of searching with information retrieval software, including Boolean logic, adjacency searching, truncation etc, and the ability to define searches by newspaper specific information such as headlines, bylines and issue dates. While taken for granted by the experienced online user the availability of this power to search large amounts of full-text data is a major advance for the user accustomed to two step access via a printed index and searching of hard copy or microfilm newspapers.

The power and sophistication of modern microcomputers and related software in the access to, and presentation of, information. Examples include the use of colour, the ability to display graphics, and the use of the Windows operating environment to allow sophisticated presentation and interaction.

To the organisation they offer: the advantages of a known fixed cost, providing many of the benefits of on-line searching without connect time, display or telecommunication charges.

A facility suitable in most cases for direct end-user access without the need for trained intermediaries or the cost of administrative systems for monitoring and, where appropriate, charging the costs of online use.

Compared to microfilm or hard copy newspapers CD-ROM offers significant space saving. A year of the contents of one national newspaper will fit on a single CD-ROM compared to 24 reels of 35mm microfilm or 12 bound volumes of hard copy. In volume terms the microfilm equivalent takes up 70 times the space of the

CD-ROM, and the bound volumes 1500 times the space. Probably no other library matches the 18 linear miles of shelving needed to house the Newspaper Library's hard copy and microfilm collections but the attractions of space saving to anyone holding substantial runs of back copies or microfilm of the titles concerned are evident.

The ability to hold information from multiple variant editions of newspapers, not just the single edition held by most libraries.

Limitations

There are, however, a number of limitations inherent in the medium which prospective subscribers need to consider.

Each newspaper title (or group of titles) is published separately, and in most cases each year of information is held on a separate disk. This means that searching of multiple years of a given title or of multiple titles is cumbersome compared to online searching (though still many times faster than searching via printed indexes and hard copy or microfilm of the newspapers themselves). Once multiple titles and years are held it is necessary to devise a system for controlling the disks and for making them accessible. Options include local area networks for multiple user access and jukeboxes or multiple drive systems for access to a number of disks at the same workstation. At Colindale we have just installed a multiple drive system which allows the user access to any one of eight different disks stored in the system by selection from a menu.

Compared to overnight updating of online files or the same day availability of original hard-copy, newspapers on CD-ROM are not up-to-date. Most of the UK titles mentioned are updated quarterly. So far only broadsheet/quality newspaper titles are available, while the date range covered is limited to the last few years. While there are obvious and valid reasons for this — their shared origins with online

files in computerised newspaper production systems, the relatively recent introduction of these, and the unsuitability of tabloid newspaper material for text only presentation — the long-term historical research use of, and interest in, newspapers is much wider than this. A recent analysis of Newspaper Library use showed only a tenth to be concerned with UK national newspapers post 1980, two-fifths of which was for those titles now available on CD-ROM.

The major collective drawback of current UK newspaper CD-ROMs is the lack of standardisation of user interface which range from the relatively simple to use to the complex and sophisticated, so that while *The Times* and *The Guardian* can both be learned and used quickly by untrained users, the interface to *The Independent* and *The Financial Times*, which both use the same 'Personal Librarian' software, is such that the first time user needs either training or to work through a tutorial to gain the knowledge to use the system successfully. In addition the type of interface determines the hardware requirements of the system on which the CD-ROMs are to be used. While *The Times*, *The Guardian* and *The Northern Echo* all work on a standard PC XT or AT type machine with 640K memory, for effective performance *The Independent* and *The Financial Times* require a 386 based machine with at least 2Mb of memory and the Windows operating environment. While the different interfaces are designed for different user groups and each has its own strengths and weaknesses the lack of standardisation is confusing and unhelpful to users having access to more than one of the products.

Information is limited largely to text only presentation, so that photographs and explanatory graphics are omitted together with the typographic and layout features of the original article which defined its content and in some cases its significance when first published.

From the point of view of historical

collections such as the Newspaper Library there are significant concerns about the archival life of CD-ROM compared to microfilm. While this may be less important to other organisations, we have worries about a medium whose physical life span may only be 10 years (2) and for which, given the rate of technological change, it may not be possible to get reading equipment in the future.

General Conclusions

Because of the limitations described it is my view that CD-ROM versions of newspapers complement rather than replace existing hard copy, microfilm or online versions and we will not use them as a direct replacement for either our hard copy or microfilm sets of newspapers concerned.

CD-ROMs do not provide a replacement for microfilm as the long-term storage and archive medium of complete newspapers, though they offer an attractive alternative to libraries which currently need to keep several years of back copies of the titles with all the problems of storage and retrieval that brings and for whose users access to the primary information content of the newspapers is the main need.

For organisations which cannot on budgetary or policy grounds provide searching of online files for users, CD-ROMs offer a fixed cost alternative, enabling the provision to end-users of the direct benefits of self-service electronic searching, while for educational institutions they allow the opportunity to offer access to valuable primary research material which in the past may have been difficult to exploit because of space considerations and the problems of subject access to the content of the hard copy or microfilm versions of the newspapers concerned.

The Future

In the months since the availability of the first UK newspaper on CD-ROM (*The*

Northern Echo in December 1990) many changes in the products have already occurred, a process which will continue. I have already mentioned the significant price cuts since the original launch of *The Times* and *The Guardian*. *The Times* has already eliminated its major weakness, the absence of adjacency searching and now offers a simple-to-use but powerful facility for this.

Greater standardisation of user interfaces must be addressed and offered, despite the competitive pressures for each product to remain different and distinct. One alternative would be for multiple interfaces to be supplied with the same database allowing the option of ease of use or increased sophistication. A technical option would be to split the interface from the data and search engine to allow users to have their own common front-end across different databases, though commercial factors may prevent this coming about.

Technical developments may lead to increased storage density, allowing multiple years of data per disk or the more systematic storage of graphics. In time this may allow the facsimile approach to the storage of newspaper information, either by the digitisation of page images or through the storage of page make-up information allowing the recreation of page images at the time of use.

The world of electronic publishing is still relatively in its infancy. Alternative electronic publishing media are certain to be developed, whether variants on CD formats or other optical or electronic storage media. Because of this and the limitations discussed above it seems likely that CD-ROM is only an intermediate stage in the electronic publishing & delivery of newspaper information. Given the choice, many users would like the power of text retrieval as seen in current CD-ROMs but to be able to see on screen and output a facsimile of the article, page and issue as originally published to produce a true electronic surrogate of the original newspapers. Such capability is a prerequisite

for the electronic publishing of the popular tabloid press for which text only presentation is inadequate. It is only a matter of time before the technology provides this kind of capability. What the timescale will be and whether libraries, other organisations and users of newspaper information will provide a market of sufficient size for such products to be developed and sustained remains to be seen.

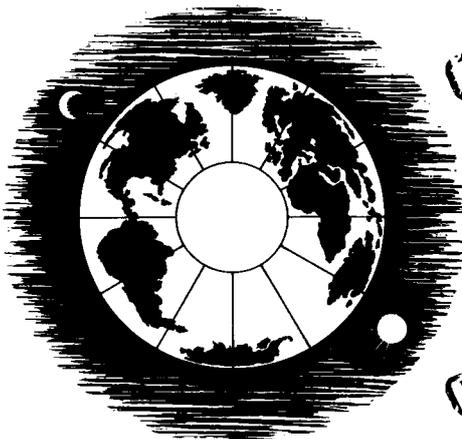
Notes and References

1. Ryan, Frank. Searching 'The Times', 'The Guardian' and 'The Independent'

on CD-ROM. *Program*, Vol 25, No. 4, October 1991, pp 319-337.

2. The archival permanence of CD-ROM and other optical disk media is unclear. Manufacturers estimate a lifetime of ten years and often more for their disks, though guarantees are usually limited. For a useful discussion of some of the problems and issues see: Marshall, Mary E. and Voedisch, Ginni. Compact disks: permanence and irretrievability may be synonymous in Libraries as well as in Roget's. *National Online Meeting, Proceedings 1990*, pp 249-254

CD-ROMs from Boston Spa



New
for 1992

Boston Spa Books is the latest addition to the series of bibliographic CD-ROMs from The British Library Document Supply Centre and contains records of over 500,000 books received since 1980.

Boston Spa Conferences, The collection of published conference proceedings held by the Centre is recognised as the largest in the world. The number of records totals 300,000, the earliest dating back to 1787.

Revised
for 1992

Boston Spa Serials contains records of over 400,000 serial titles together with details of the holdings of the Centre.

All of the Centre's collections of serials, books, proceedings, reports, translations, theses and official publications are available through the document supply services.

For further information contact:

Mike Curston,
CD-ROM Development
The British Library Document Supply Centre,
Boston Spa, Wetherby,
West Yorkshire LS23 7BQ, United Kingdom
Tel: 0937 546061 Fax: 0937 546333 Telex 557381