DIGITISATION OF NEWSPAPERS AT THE BRITISH LIBRARY

by

Edmund King

Abstract

The newspaper collections of the British Library are among the world’s finest. The repository at Colindale, North London, UK, holds some 20 miles of UK, Colonial, and overseas newspapers published since 1800. The author charts the work of the British Library over the last dozen years, in using the technology of the time to explore how digitisation can provide greater access to portions of the British Library’s newspaper collections. Software and hardware developments have quickened since 2000, and projects have been completed on a small scale to prove the concept of internet access to texts using software which can enhance readability of texts, and also can search all of the texts, with results immediately available to users. In currently undertaking a large project to scan up to 2 million pages of nineteenth century newspapers, the Library is moving to a larger level of scale in its programme to provide wider access to its newspaper collections.

Keywords: British Library; historic newspaper collections; digitisation; Burney Newspapers; Collect Britain; Penny Illustrated Paper; British Newspapers 1800-1900 Project.

Acknowledgement: The author is grateful to his colleagues in the British Library—John Tuck, Kristian Jensen and John Goldfinch—who read the text and made suggestions for its improvement.
DIGITISATION OF NEWSPAPERS AT THE BRITISH LIBRARY

Introduction

It can be difficult for many to grasp the scale of newspaper publishing in the United Kingdom. Taken as a whole, the huge production of newspapers since, say, 1700, provides an enormous resource for research on all subjects for all of the UK, both urban and rural. Newspapers have a long history of being copied after publication. For decades, even hundreds of years after their publication, researchers of all kinds, all over the world, turn to newspapers for information relating to their needs. For those libraries that have collected newspapers (particularly national libraries), the need to provide access to newspaper texts has posed a dilemma, given the often very poor nature of the paper the newspaper was printed on. The great need to prevent undue wear and (literally) tear upon the paper has provided impetus to copying texts, to allow continuous public access.

The British Museum Library, and, since 1972, the British Library have systematically collected newspapers through legal deposit arrangements, purchase and exchange. Some facts and figures relating to the British Library newspaper collections demonstrate the scale of this historical activity. British Library Newspapers is the UK national archive collection of British and overseas newspapers; it is a large, integrated newspaper service, combining facilities for the collection, preservation, and use of newspapers all on one site. The collections, which also include popular magazines and periodicals, are made available in hard copy, in microform, and on CD-ROM in the Newspaper Reading Rooms in Colindale, northwest London. The collections housed there consist of:

- over 664,000 bound volumes and parcels, occupying some 32 kilometres (or 20 miles) of shelves;
- over 370,000 reels of microfilm, on 13 kilometres (or 8 miles) of shelf space;
- holdings altogether occupying some 45 kilometres (or 28 miles) of shelving in total

Over 52,000 separate newspaper, journal, and periodical titles are currently held in British Library Newspapers, Colindale.

Microfilms of Newspapers

The need to copy newspaper texts is obvious to all – newsprint is a poor quality paper whose longevity is not assured. Newspapers also go out of print very quickly, unlike books, which may remain in print for some time after their publication. In the last fifty years, the preferred medium for copying newspapers has been microfilm. It is an advantage to all of us that the photography techniques upon which microfilm is based have been with us since the early 1840s. The durability of the photography is well established, provided quality procedures are followed in the creation of the film, its processing and copying, and the storage of all copies of the microfilm. When these
procedures are properly followed, microfilm negatives are likely to last for over 200 years, with the newspaper texts on it being available for many generations, as positive copies.

Of course, microfilm has its limitations, some of which also apply to the original paper: only one person can use a given film at a time; users have to load the film in reels onto reading machines; pages have to be read sequentially (or, the skipping of pages has to be done in sequence); the entire contents of an article, or of a page, has to be read, for the user to avoid missing relevant information; copies of an article, a photograph, or of a page have to be paid for.

**Digitisation of Newspapers: Underlying Factors—and Problems**

A practical distinction has emerged in recent years between current newspapers which are “born digital”, and those created in the letterpress era, which require conversion to digital formats using scanning equipment. Only in the last ten years or so has it been possible for libraries to consider alternatives to consulting the microfilm of the back runs of newspapers. It is possible now to plan to have microfilms scanned digitally, with the digital images capable of further manipulation by software, to permit a good degree of readability, with the texts also being searchable. This situation has been brought about by:

- a huge increase in computing power, combined with miniaturisation of parts, available to millions via desktop and laptop computers
- a great fall in the prices of personal computers brought about by mass production
- the arrival of the world wide web
- the rapid development of software to link computers, together with sophisticated software to provide information to users
- the increase in sophistication for the indexing of all kinds of digital information, to ensure its availability to many other users
- the development of hardware and software to create pictures digitally, store them, and transmit them to other computers

Since 1999, the pace of change has been rapid. It has become possible to have information indexed from all kinds of sources. This step is normally undertaken after optical character recognition processes have been applied. Huge efforts have been invested in developing software to do this automatically, rather than link searches for text via keywords typed manually. Texts created digitally using a computer have been the simplest for the application of such software. On the whole, texts in manuscript or in letterpress have been more difficult for software to work with, as the variables of character recognition are so much greater.

As with many other letterpress texts, newspapers pose problems for scanning and optical character recognition (OCR). Some of them are:

- text characters may be over- or under-inked
- text characters may be worn and not print properly
• there may be “bleed through” of ink from either side of a sheet of paper (this occurs quite frequently for pages which have wood engraved illustrations surrounding text, or where the heavier inking needed to reproduce the illustration results in bleed through to recto/or verso of a sheet)
• paper may be creased, resulting in loss of text
• the spine curvature of bound volumes may make text difficult for scanning machines to recognise, and also for OCR
• paper may be cockled
• significant variation in font sizes may be found within individual issues of a newspaper
• layout of text into columns is not always regular, and the layout could vary significantly over time

These problems are compounded by changes in the frequency of publication, the size of paper, the titles of publications, and the placement of subject matter within daily or weekly issues of newspapers (e.g., parliamentary reports, law/crime reports, foreign news, sports reports).

Indexing

Offering scanned images without any fully searchable text would only partially fulfil users’ needs, as many (if not all) desire to search all of the texts available to them. It is necessary to consider providing navigation within the digital data through the use of indexing software. Here the structure and layout of eighteenth and nineteenth-century newspapers present a considerable challenge, largely owing to the many differences in layout of the newspapers published in these 200 years.

Most users will require access to a topic by date or a range of dates; this is obviously quite easy to provide. Others may also require access by title, place of publication, publisher, and issue number.

Still other users will require access by personal names or keyword search terms. Apart from general background information, searching could range from specific reports of births, deaths and marriages, through legal cases to advertisements of all kinds. Often the information contained in newspaper reports is unique to those sources. At present, manual indexing is laborious and time consuming. Of the UK national newspapers, only the Times is indexed fully, and to achieve this took many years.

For other researchers, a subject approach would be more appropriate. However, unlike modern newspapers, there are less usefully descriptive “headlines” to the various reports. Headings vary greatly for different newspapers. They can vary to quite a degree within a long run of one newspaper. Any attempt to provide a usable subject index may therefore need to be based on a degree of manual editorial intervention, which would require additional resources.
In order to set up a viable project in the digital context, it may be necessary to limit the indexed data to title and imprint information, supported by the improved OCR facility that now is available. In the longer term, subject indexing to enable keyword searching remains a desideratum.

Recent Digitisation Developments outside British Library

Today literally thousands of newspapers provide access to all or part of their content on the world wide web. There are also portals on the web that offer access by country, geographical region, or title. For example, the Nettizen.com Online Newspaper Directory site offers access by title to the top world newspapers, as well as access to newspapers by each world region. Top 100 Media, on the other hand, provides, as its name implies, access to 100 online news resources. It is only when one enters the site for each individual title that one sees whether the information is offered free or via subscription. It would appear that on-screen advertising pays to enable the display of much text on sites being offered free of charge. Some national libraries are also creating portals of their own; for instance, the National Library of Australia’s Australian Newspapers Online portal provides access to many of that country’s newspapers by title, state, and town.

Through digitisation, many older newspapers can now also be found on the world wide web. A number of national libraries are directing online users not only to their catalogues of newspapers or to descriptions of their newspaper collections, but to descriptions of their digitisation projects as well; for example:

Library of Congress

The library’s Newspaper and Current Periodical Reading Room has an online guide to its newspaper collections. The library also provides information on the United States Newspaper Program, probably one of the largest preservation projects undertaken to microfilm older U.S. newspaper titles, state by state, since the early 1980s. It also has information regarding the National Digital Newspaper Program (NDNP). The potential for scanning the microfilm exists; funds are yet to be found to begin a project to do this. The National Endowment for the Humanities (NEH) is planning to facilitate digitisation by funding the NDNP. Ultimately, over a period of approximately 20 years, NDNP will create a national digital resource of historically significant newspapers from all the states and U.S. territories published between 1836 and 1922. This searchable database will be permanently maintained at the Library of Congress (LC) and be freely accessible via the internet. LC will also digitize and contribute to the NDNP database a significant number of newspaper pages drawn from its own collections during the course of this partnership with the NEH.

National Library of New Zealand

The Papers Past Project shows selected nineteenth century newspapers and periodicals, with digital images of over 400,000 pages from 25 publications being available.
National Library of Canada

The library has created an Electronic Collection; it intends to create digital images for newspaper texts as part of this initiative. 8

National Library of France

The Gallica 2000 project has created a substantial body of texts in digital form available for searching. Newspapers do not appear to feature in the list of texts available. It is likely that full digital images of French newspapers will be created and put on the site in due course. 9

National Library of Austria

ANNO is the library’s virtual newspaper reading room, offering access to a growing number of digitised newspapers. 10

National Library of the Netherlands

The War & Revolution. Digitisation from Microfilm from Dutch Newspapers 1910-1919 project is conducted in the framework of the Dutch national programme for preservation of library material, Metamorfose, which is an initiative of the Netherlands Ministry of Education, Culture and Science. The selected part of the collection contains historic newspapers from the period 1910-1919. 11

TIDEN – A Nordic Digital Newspaper Library

TIDEN is a collaborative project of the Helsinki University Library, the National Library of Sweden, Statsbiblioteket - the Danish State and University Library in Aarhus, and the National Library of Norway for the digitisation of newspapers on microfilm. 12

British Library Digitisation Strategy since 2001 13

Alongside these national library developments, the British Library sees digitisation as a way to:

- maximise use of the collections by facilitating a greater volume of networked access, both in the reading room and remotely, and by providing the enhanced functionality intrinsic to the digitised items
- reflect national and international priorities for wider access to, and enhanced use of, integrated collections of digitised educational, cultural or scientific materials which have a combined value greater than their component parts
- contribute to the conservation of original analogue materials for future users by substituting digital surrogates
- preserve the collection of analogue sound recordings
• generate income from those products with market appeal that can be exploited commercially by a partner, or the British Library itself, consistent with the aim of maximising accessibility to the collection

There have been a number of digitisation projects initiated recently by the British Library. Some recent examples include:
• Turning the Pages
• Shakespeare Quartos
• Gutenberg Bible

Newspaper Digitisation Projects in the British Library

Burney Collection of Newspapers
The Burney Collection of Newspapers, one of the British Library's most important and heavily used microfilm collections, was chosen to provide materials for the library's experimental microfilm digitisation project. This collection begins with what are actually Parliamentary papers from 1603; the earliest items which are recognisable as newspapers date from the early 1620s. It is in eighteenth-century London newspapers, however, that the collection is particularly rich: all the major titles are included, such as the Daily Courant from 1702 to 1735, the first daily newspaper published in London, and the London Gazette from 1665. Periodicals are also included, such as Addison and Steele's famous Tatler (1709-1711) and their Spectator (1711-1712).

Also represented are English provincial titles from 1712, such as the Stamford Mercury of 1728, Irish newspapers (the earliest being the Dublin Intelligence of 1691), Scottish titles from 1708 onwards, and many eighteenth-century American ones too, including the New England Courant (1721-1723), on which Benjamin Franklin (1706-1790) worked, before he moved from Boston to Rhode Island after having published the paper's final issue on 25 June 1726.

The microfilm sets of the Burney Collection seemed an ideal test case for this project. In addition to the intrinsic importance of the collection and the poor quality of the originals and microfilm surrogates, those setting up the project relished the challenge represented by eighteenth-century printed text, particularly by rare or obsolete characters such as the long “s”.

In 1992, the library acquired a Mekel M4000XL microfilm digitiser. In the following year, a project was set up to test the equipment as a part of the library’s wider Initiatives for Access programme. Other principal aims of this project included:
• to examine the most appropriate storage medium for digitised material
• to examine the implications for staff resources and operations
• to determine the technical and procedural implications of supplying material
digitised from microfilm to readers in reading rooms
• to investigate microfilming standards and their importance for microfilm
digitisation

In the 1990s it was estimated that there were some 650,000 frames of microfilm
representing pages in the Burney Collection. The project decided to focus on scanning a
particular period of the eighteenth century, and selected the early years of the French
Revolution (1789-91), completing this work early in 1995. Numerous problems were
encountered during the process, many of them owing to the nature of the original
material. The library was also reluctant to use the archival negative films in case these
preservation masters were damaged in the process. It was discovered that scanning
parameters needed to be reset almost for every frame in order to secure acceptable
results. Clearly this affected productivity quite seriously and early estimates suggested
that it would take eleven years to scan the whole collection by this process.

By upgrading the equipment after March 1995, continuous scanning was made possible
with obvious benefits for the production process. The estimated time to scan the
collection was reduced to some eighteen months. Indexing the data raised as many
problems as scanning. None of the current OCR packages then available produced
acceptable results with eighteenth-century printed texts. Within the newspapers
themselves, there were few obvious index points (such as ‘headlines’). Manual
intervention was necessary to create free text or subject indexing.

By March 1996, when the project closed, some 21 gigabytes of data had been collected
and a productivity rate of 6,000 frames per month achieved. The results showed that
digitisation of microfilm of early printed texts could produce an acceptable digital
surrogate. But, without viable indexing tools, that surrogate would be impossible to
access or use.

By the late 1990s, technical developments led to initial discussions with potential
partners and suggested the technical and commercial viability of this project. Tests were
carried out on sample runs of newspapers from various points in the eighteenth century.
The film of the entire collection was digitised on a Mekel M500 greyscale production
scanner.

At this time, further funding is being sought by the library to process the scanned images
and to make the all of the texts available for searching via a website.

The British Library Newspaper Pilot 15
In the first half of 2001, the British Library and other partners 16 collaborated to produce a
prototype system for the digitisation, indexing, and presentation of historic newspapers
from the newspaper collections of the British Library. The overall goal of the project was
to allow for online accessibility to the library’s historic newspapers, and in particular, those of the nineteenth century.

The goals of the project were
- to digitise a critical mass of microfilmed newspapers in a relatively short space of time
- to index the digitised documents automatically
- to deliver the documents in a sophisticated online format
- to provide advanced searchability of the texts
- to present the results to the widest possible audience for feedback

There were a number of problems inherent in the nature of older newspapers which needed further testing by the project:
- complex page layout
- radical differences in layout over time
- titles with very large letters
- narrow spaces between lines of text
- poor microfilm image quality
- skewing, or ‘curvature’ of text owing to tightly bound volumes of newspapers
- ‘noise’ resulting from dirt on the original paper, from dirt on the scanner lens, etc.
- broken vertical and horizontal lines between articles, advertisements
- broken font characters

The project decided to utilise the Olive Software system, which deployed the ActivePaper Archive components.

Out-of-copyright newspapers were selected for the pilot because the timescales to which the team were working meant that protracted negotiations with publishers would slow the process unacceptably. The software needed to be tested for its ability to handle such things as imperfect printing, multiple size fonts on a single page, and bleed-through of print from the recto to the verso of pages.

Eighteen reels of duplicate negative microfilm were selected, containing newspaper issues which focused upon British ‘national’ events. Full runs of the same few months in each year were chosen for each newspaper title. The intention was for users to look at a variety of perspectives on the same events or issues. Microfilms of a range of ages or states were duplicated, the aim being to test the technical issues attendant upon the digitisation of microfilm as well as the acceptability of the content. In the first stage, microfilm reels were scanned to 300dpi TIFF images by OCLC Preservation Resources. The images were then shipped on CD to Olive Software’s processing facility.

The digitisation was divided into two parts—precision scanning and image processing—with a digital newspaper archive being generated from the image processing. The ActivePaper Archive, created by Olive Software, used an image processing technique called ‘segmentation’, which breaks each page of newspaper
text down into its smaller information units (articles, pictures, advertisements). The
software identifies them. By using artificial intelligence and a patented fuzzy search
technology, the software attempts to overcome the obstacles of poor image quality
and complex page layout very common in older microfilmed newspapers.

For searchability, ActivePaper Archive relies on OCR-generated word patterns,
stored in Extensible Markup Language (XML) format. The software uses APFS
(Adaptive Predictive Fuzzy Search, patent pending), a fuzzy logic search technology,
to compensate for text inaccuracies by applying fuzzy logic according to the
probability for error in each word pattern.

In ActivePaper Archive, newspapers and documents run through the image
processing stage are converted to Activepaper XML. Traditionally, XML holds text
and its structure, but ActivePaper Archive goes further by tying the XML to images.
The product uses three XML layers—one based on the NewsML/NITF standards, one
on the Dublin Core, and a third on PRML, or Preservation Markup Language. PRML
maps the newspaper’s layout, recording coordinates for each piece of text and each
page object.

The system and the data were created and have been live since 2001. Usage is around
25,000 user sessions per month, and this is without major promotion and publicity.
Responses to the pilot proved wholly positive. It is estimated that some 20,000 pages
of newspaper pages were processed in two months, containing some 500,000 articles
and advertisements. It has been widely realised that the use of software similar to the
Olive Software product, via rapid and comprehensive searching, could transform the
study of modern history, cultural studies, and other academic disciplines by unlocking
information about all aspects of life that were printed in newspapers.

Collect Britain Project
The aim behind this project was to make portions of the British Library’s collections
much more widely available. Images and sounds from the British Library's collections are
intended to evoke memories of locations in Britain and beyond. Maps, views, prints and
drawings, early photographs, advertising ephemera, dialect recordings, the whole run of
the Victorian *Penny Illustrated Paper*—these are all among the 90,000 items presented
within this website.

The British Library holds the only complete run of the *Penny Illustrated Paper*, which
provides a valuable and entertaining source of detailed information on everyday life and
historical events in Britain and across the Empire. It was decided to include all of the text
of this paper, published weekly between 1861 and 1913, as one part of the whole
project.
Today, mass-circulation newspapers, crammed with images of people and dramatic events, are taken for granted. In the nineteenth century, illustrated newspapers were excitingly new—and, as now, influential. The Illustrated London News led the way when its first weekly issue went on sale in May 1842. By 1863, the circulation of The Illustrated London News reached an impressive 300,000 copies a week with a cover price of six pence. Other publishers were not slow in setting out to capture their share of this expanding market. The Penny Illustrated Paper followed the lead of The Illustrated London News in offering articles covering a wide variety of subjects and generously illustrated with woodcuts. The size and content of the illustrations varied greatly, reflecting the degree of popular interest in current events and the ‘celebrities’ of Victorian life.

Ebenezer Farrington published the first issue of his weekly Penny Illustrated Paper on 12 October 1861. The title proudly proclaimed itself in capital letters and carried the motto: ‘With all the news of the week’. The newspaper adopted a campaigning stance, from the start: "In producing an Illustrated Paper for the million, let us plainly say, we want it to be esteemed the friend of the million," read the first editorial. Its aspiration was to use the cheap press to help solve "...the terrible amount of suffering, disorder, and vice that must be dealt with in a more vigorous fashion than hitherto..."

Contents were divided by sub-titles, printed in gothic lettering. For example, the edition for 24 January 1863 had Topics of the Week; Home News; Accidents, Inquests, etc; Guy Waterman’s Maze (fiction); Picked Up at Sea; Sporting News; Recreations; Foreign News; Law & Police; Gossip; and Advertisements.

One of the predominant features of the Penny Illustrated Paper was the honesty of the reporting, together with the clarity of its expression. The sometimes crude, yet vividly depicted, illustrations gave an immediate impression of events occurring in that week. Looking at the results today, we cannot fail to appreciate the achievements of the editors in not only gathering wide-ranging reports each week, but also ensuring that a substantial number of woodcut engravings, and later photographs, were organised in time to go to press.

For over fifty years, every week, this work continued alongside competitors whose weekly cover price was much greater. During the closing years of the nineteenth century, the rise of mass-circulation national daily newspapers provided even more competition. In January 1908, the newspaper’s title was changed to P.I.P., Penny Illustrated Paper. Five years later, the title was lost altogether when it was absorbed by London Life.

The inclusion of this material marked another stage in the evolution of software, as the majority of the texts of this newspaper are indexed, and capable of retrieval by users who have internet access.

**British Newspapers 1800-1900**
In 2004, the British Library secured funding from the Joint Information Systems Committee, JISC, as part of its Digitisation Programme to film and scan up to 2 million pages of nineteenth century newspapers.

The project, reflecting the British Library's vision of 'helping people to advance knowledge to enrich lives', will offer users in Higher Education (HE) and Further Education (FE):

- Free access via a sophisticated searching and browsing interface on the web to a virtual library of nationally, regionally and locally important digitised British newspapers from 1800 to 1900
- Free access to a broad range of valuable learning materials, many of them unique, and hitherto available only in London-based reading rooms, in hard copy or in microform
- The ability to search across the different newspaper titles to draw together materials relating to a wide range of research and learning topics
- Content focusing on London national newspapers, English regional newspapers, home country newspapers from Scotland, Wales and Northern Ireland, and titles in specialist areas such as Victorian radicalism and Chartism

The project has defined a number of keys to success:

- Close involvement with HE and FE representation on the Project Board in the overall management of the project
- The early identification of the final content to be digitised based on legal advice and through the user representatives on the Project Board
- The selection of a high quality supplier through a tightly managed procurement process
- The establishment of a strong management team

The British Newspapers 1800-1900 project relies on a combination of in-house British Library work and components provided by third parties. The allocation of work between in-house operations and third parties is based on where value can best be achieved, balancing the cost effectiveness of competitive tender with the optimum deployment of experience and expertise from the library.

The aim is to select and to digitise complete runs of UK newspaper titles that were published between 1800 and 1900. It is proposed to select a mixture of UK national, regional and local newspaper titles, whether microfilmed or not, which reflect the social and political developments of the times in which they were published. The selection will be to mix currently well known national and regional titles with those titles which are now less well known.
Surveys of recent UK newspaper users confirm that the library’s readers use newspapers to support a wide variety of research projects reflecting the great diversity of subject matter contained in newspapers. This is especially the case with the UK London national newspapers and the overseas newspapers. The surveys also indicate that a significant number of our readers use a mix of newspaper categories and do not necessarily restrict their research to one particular category of newspaper or type of content.

The library's strategy for the successful completion of the project comprises in-house activities for material selection, conservation and microfilming undertaken by library staff and the procurement of a solution through competitive tender for digitisation, content extraction and web delivery. The project is being managed by British Library staff reporting to the Project Manager and the Project Board. The project’s website is to be updated at regular intervals.

**Functions in the “Digital Chain”**

All of these British Library projects for newspapers, in their different ways, have moved and are moving the library further forward along the path of opening up its huge newspaper collections. At this point, there are a number of functions in securing full public availability for the newspaper texts that need to be considered:

- Proper preparation of the original newspapers, in sequence, for either microfilming or scanning
- Effective recording of the ‘metadata’ of the newspapers (essential for links to digital images)
- High quality microfilming of the newspapers
- Quality scanning of the microfilm or the originals
- Quality assurance of the scanned images
- High quality Optical Character Recognition of the scanned images
- Some form of division of the newspaper pages into ‘zones’ or ‘segments’
- XML capture to provide the essential link to the texts
- Use of software to index individual articles by subject words
- Use of software to enable searching of the entire set of texts; the searching to provide a high level of success, via the use of probability software to determine the correctness of characters and of words
- Use of software to provide a ‘host’ for the data, and to permit users to access it via the internet
- Ensuring that the data is preserved for future use

All of us are aware now that these steps can be carried out at a certain production cost, and that the costs can be related to a budget for the work that is planned. It will be for individual project planners to determine which of the above can be paid for.

**Conclusion**
The past few years has seen a significant expansion of activity in making copies of older runs of newspapers available via the internet. It is anticipated that as commercial and public body investment grows, far more will be available soon. It is also likely that there will be far greater ease of access to texts available as a result of the application of sophisticated software. This will significantly increase the ability of researchers to search quickly through huge quantities of printed newspapers, which will be a vast improvement over the former manual process. The aim is to make the enormous amount of information contained in British Library newspapers, covering all subjects imaginable, more accessible to all. This prospect is all the nearer due to recent developments, and the effects of such information being openly available will be far-reaching.
Notes and References

1 See: http://www.nettizen.com/newspaper/ (site visited 10.6.2003)


10 See: http://anno.onb.ac.at/ (viewed 4.11.2004)


12 See: http://tiden.kb.se/ (visited 18.10.2004)

13 See: http://www.bl.uk/about/policies/digital.html


16 Marilyn Deegan et al. The Digitisation of Historic Newspapers; The British Library Newspaper Pilot, pp.71-87. See also the data at: http://www.uk.olivesoftware.com
The other partners in this project were: OCLC, Olive Software, the Malibu Hybrid Library Project at King’s College London, and the Forced Migration Online team at the Refugee Studies Centre, Oxford University.


