THE BURNEY NEWSPAPER DIGITIZATION PROJECT

By Dr. Henry Snyder, Director of the Center for Bibliographical Studies & Research
University of California at Riverside

In public libraries throughout the Western World the newspaper is the most widely used printed medium in their collections. For scholars or the general public they are an invaluable though intractable resource, frustrating to use because of the lack of indexing, the unwieldy format, whether microfilm or hard copy, the fragmentary and arbitrary survival and their scattered locations in a vast range of repositories. For the Early Modern Period their excessive repositories. For the Early Modern Period their excessive rarity and fragility increases difficulty of access and use. But this access is undergoing a technological revolution through the introduction of digitization and distillation (optical character reading).

The Center for Bibliographical Studies at the University of California, Riverside in partnership with the British Library is employing these new technologies to open up the full resources of the Burney Collection, the great collection of eighteenth century English newspapers. Two successive NSF grants have enabled the partners to complete the digitization phase. The number of frames digitized totals 988,579. We are now planning the second distillation phase. The Burney Collection is the most important in the world for the eighteenth century London newspaper. During this period, the first daily was introduced, the first provincial newspapers outside the capital were begun, and the format, design, type size, and features of what we are familiar with in today’s newspapers all evolved. Because of their fragility access to the originals is denied. Access is only possible through a film copy available in the reading room at St. Pancras. A commercial version distributed by Gale has both additions and omissions. As part of the digitization process an index by issue was created reflecting title, date of publication, volume and issue number and Burney location.

This index in turn will be used to access the set when it is mounted on the web. Heretofore the vagaries of hand set type, inequality of inking, wear and discolouring to the originals made optical character reading impractical because of the low accuracy rate. New advances in technology combined with manual intervention to provide an accurate index to article and heading titles, authors, keywords, etc. now make distillation technically feasible. By creating a text-searchable product, access will be revolutionized. There are no existing indexes for the newspapers. To ferret out information requires painstaking page-by-page reading and when the total number of pages approaches a million, that can be a near impossible task. The new technology will permit Boolean, keyword searching. When an article is retrieved it will be reproduced either by itself or as part of the page.

The British Library has received an award from the Joint Information Sciences Committee (JISC) to digitize two million pages of nineteenth century newspaper pages. It is the plan of the partners to create compatible files so that a single search engine can manage the two. As scholarship does not recognize century limits the value of the Burney file will be greatly increased by its association with nineteenth century material. It is the intent of the partners to make the Burney Collection available free to users over the Internet. But this will only be possible if the balance of funds required can be secured from a public, non-profit agency. Otherwise recourse must be had to a commercial sponsor, which would necessitate a use charge for readers.

A digitized, text-searchable version of the Burney Collection will be an exemplar of the latest technology and a model for other newspaper digitization projects throughout the world. It will serve a limitless range of students and scholars working in the eighteenth century, not only in British studies, as the newspapers published stories of international interest. Because of its free access and the infinite range of subjects covered, it will be used by K-12 students as well as senior scholars, by literary and historical scholars as well as amateur genealogists.
As with last year’s meetings, the business meeting was held on the first day followed by the country reports as part of a larger seminar on newspaper preservation. Our observers at the business meeting from the Shanghai Library were Peijuan Wu, Feng Beng, Min Wu and Yijing Zhang. Mr. Huizu Wu was instrumental in organizing the meetings and arranging the visits in Shanghai.

Harmut Walravens reported on the 2003 IFLA Berlin Open Programme (114) Newspapers and Copyright Developments, in particular on the recent European legislation on copyright and the difficulties in securing consent to digitize newspapers, the licensing rights going from newspaper publishers to third party businesses and the resulting denial of free access via libraries to newspaper content, the lack of knowledge surrounding the ownership of microfilm and the rights of third party companies to microfilm, the copyright rights of individual journalists. He also mentioned that the law supports the mission of National Libraries to be the future repositories for newspapers received in digital form.

In 2005, IFLA in Norway’s theme for the Open Session will be Digital Newspaper Projects and the session will be a joint one with the IFLA Serials Section. Prior to the Open Session, there will be a visit to Mo I Rana, the National Library of Norway’s Digital Preservation storage facility near the Arctic Circle for the Newspapers Section and Preservation and Conservation Section. More details will be available in the next newsletter.

More work needs to be done to achieve funding for a Newspapers Management Handbook and members will revise the proposal to include a benefits statement and sample pages. The project is considered to be worthwhile continuing to pursue, as there are few if any modern publications, which can be used to train librarians dealing with newspaper collections, particularly in Africa.

The membership of the Section stands at under 30 and we hope to raise this to 50 by encouraging at least one other library per country to join.

An international list of newspapers is available at the Library and Archives Canada web site at http://www.collectionscanada.ca/8/16/r16-211-e.html (English) and http://www.collectionscanada.ca/8/16/r16-211-f.html (French). Members agreed that this site could be mounted on to IFLANET and that it could also be customized and

REPORT OF THE 17TH BUSINESS MEETING
(4th Business Meeting, Newspapers Section, IFLA)
held at the Shanghai Library, Shanghai, China on 29th and 30th March 2004

Speakers:
Alfonso Quintero (Secretaria Ejecutiva de ABINIA, Caracas)
Proyecto «Digitalizacion de la Prensa Latinoamericana del Siglo XIX» (Digitalización de la Prensa de la Gran Colombia : 1820 a 1830)

Jorge Orlando Melo (Biblioteca Luis Angel Arango, Bogota)
El periodismo colombiano antes de 1900 : colecciones, microfilaciones y digitalizaciones

Adán Benavides (Benson Latin American Collection, University of Texas, USA)

Yesterday’s News Today: Preserving Mexican Newspapers, 1807-1929

This will be the first IFLA conference to be held in South America. Henry Snyder has kindly arranged for the speakers for this session.
updated to include more newspaper sites as well as perhaps mirrored on a university website in a country where the Newspaper Section holds a future business meeting.

At the end of the meeting, the Director of the Shanghai Library, Dr. Wu Jianzhong presented commemorative plates of the Shanghai Library to the Section members.

On the second day, the Section members attended the one-day conference on Opportunities for Newspaper Access and Preservation in China and presented the country reports. Consecutive translation was available and 60 librarians including librarians from the Shanghai and branch libraries and university libraries as well as information professionals in the field of newspaper digitization attended. and papers were given on the following topics: International Newspaper Work by Dr. Hartmut Walravens, Director of the Department of Bibliographic Services, Staatsbibliothek Berlin, Newspapers in the Shanghai Library by Min Wu, Shanghai Library Reader’s Service Centre, UK NEWSPLAN Project: History, Present and Future by Edmund King, Head of the British Library Newspaper Library, Newspapers in the Royal Library by Lars Olsens, Head of the Newspaper Section, The Royal Library National Library of Sweden, Newspaper Access and Preservation in France in 2004 by Isabelle Rollet, Head of the Newspaper Section, Bibliothèque nationale de France, Digitising Newspapers: Advancement in Production and Use by Majlis Bremer-Laamanen, Head of the Preservation and Digitization Centre, Helsinki University Library—The National Library in Finland, Introduction on The Digitising Newspaper Works of China by Dr. Yuzhi Zhang, General Manager of the Datum Data Company Limited, Making Known a National Collection of News Media in Canada by Sandra Burrows, Newspaper Specialist, Reference and Information Services, Library and Archives Canada, Newspaper Resources in College Libraries: collection and services by Dr. Keqian Xu, Director of Nanjing Normal University Library, British Library Newspaper Collections: Development and Future Access by Geoffrey Hamilton, IFLA Newspaper Section Member, Development and Utilization of Newspapers in District Libraries of Shanghai by Ying Si, Director of the Shanghai Putuo District Library, and The legal deposit of Internet Newspapers in Norway by Jonny Edvardsen, Director of Collections and Bibliographic Services, National Library of Norway. Attendees had the opportunity to participate in guided tours of three Shanghai libraries following the meetings: Huangpu, Jiaotong University and the Bibliotheca Zi-Ka-Wei (Xujiahui). The Huangpu District Library no. 2 (Nanshi District Library) joined the Shanghai Library in 1991. Of particular interest to the IFLA Newspaper Section members were the 500 newspapers and peri-
odicals, the clipping service on world travel and culture and the childrens’ periodicals reading room. The Jiaotong University houses over 4800 Chinese and foreign periodicals and newspapers. The library’s focus is mainly in the fields of Communication, the Physical Sciences, Engineering, Electromechanical Sciences, Information Management and Agricultural Science and it provides 500 personal computers on the three library campuses. We visited a beautiful older library also, the Bibliotheca Zi-Ka-Wei which was formerly part of a Jesuit mission and founded in the Xujiahui district in 1849. The current library houses many rare books and newspapers and the bookcases are painted with scarlet lacquer. The western part is multilevel with iron columns around the room.

IFLASURVEYON HOLDINGS OFAFRICAN NEWSPAPERS KEPT IN ARCHIVES AND LIBRARIES INAFRICA – UPDATE

By Else Delaunay, IFLA Newspaper Section, Bibliothèque nationale de France, Paris 2004

In June 2003, the IFLA Section on Newspapers together with the IFLA Preservation and Conservation Section took up again the JIPCA survey on the safeguard of newspapers and periodicals in Africa which was conducted by the IFLA PAC in 2001. The Section has limited the new Survey to newspaper only, published in Africa (retrospective holdings as well as current files). The newspaper is by definition an ephemeral item which is normally printed on poor oversized paper with poor ink and its preservation is precarious. In most cases holdings need restoration and/or transfer to another medium. In this regard, the Section wants to locate and know about newspaper holdings kept in Africa in national or university libraries as well as in national archives.

General Results

The Survey has been going on from July 2003 to January 31, 2004 and the last replies were received as late as the end of February. 125 institutions in 50 countries were contacted of which 43 (34.40%) in 28 countries replied (56%). When the email address and the name of the person in charge of the newspaper collection were available, the replies were quicker and of a higher level. Complementary explanations were possible in a few cases. In most cases, the fax number available was not the correct one and it became necessary to send the fax by post. This explained the late development of the survey results. The results also showed up some confusion in the interpretation of the questions which meant a «no reply» to some of the questions. The same confusion was found in the list of titles adjoined to the Query forms (27): many of these lists included periodical titles in addition to the newspaper titles (dailies and weeklies) and even some non-African titles. However, on the whole, the replies allow us to get an idea of the newspaper holdings in the different institutions. Lack of funding recurs generally and rather precise information is also given about the immediate needs of some libraries and archives. Several institutions have no microfilms and digitized items and access to the Internet is not available. Many African colleagues prefer digitized newspapers, especially because of bad climatic conditions for storage. Nevertheless the microfilm is still the best way of long-term preservation of the documents.

Survey Follow-up

According to the letters sent together with the replies, it is obvious that our African colleagues expect the Newspapers and PAC Sections of IFLA to take efficient action quickly as most of them have great difficulties in preserving, collecting and giving access to their newspaper holdings. A second simplified but precise questionnaire concerning two priority newspaper titles for each institution
will be circulated in June 2004, firstly, to the institutions, which mentioned immediate potentially disastrous loss of materials. The deadline will be September 30, 2004. The estimates for the need of reading and storage equipment must be calculated very precisely. A covering letter will be circulated with the Query Form. At the same time, the list of titles will be examined in order to search for titles, which are already available on microfilm. If this is the case, a positive duplicate of the film will be proposed. Replies to the second questionnaire will allow the Newspapers and PAC Sections of IFLA to prepare a case for funding support from potential private and public sponsors.

The African Digital Library Centre

The African Digital Library Project is maintained by Dynel Patel at the University of Cape Town in partnership with the New Zealand Digital Library Project. The web site provides several document collections, including historical documents, humanitarian and development information, computer science technical reports and bibliographies, literary works and magazines. All are accessed through searching and browsing interfaces provided by the Greenstone Digital Library Software.

This software allows users to create collections of any digital data (including images, audio, video, text, x-rays, paintings etc.) and allows them to publish these collections on the Internet or on a CDROM. Collections can be searched in a number of ways. Collections that are exported to a CDROM are completely self sufficient and this allow users to browse collections without the need of Internet access. This software has been designed specifically for developing nations and caters to all operating systems (WINDOWS, UNIX, MAC etc.) at any level. It also provides multilingual support both for the interface and content and does not require powerful hardware. Extensive testing of the product has been done by UNESCO (who is also a co-founder of its development stage). The software is free of charge under the GPL license. More information is available at: http://greenstone.cs.uct.ac.za/gsdl/cgi-bin/library or http://www.greenstone.org/cgi-bin/library

Information on the Greenstone Digital Library Software was taken from the website listed above.

Dutch Newspapers 1618-1700: A Pilot Project as Part of the Digital Library of Dutch Newspapers 1618-1869

In the history of printed news, the Dutch Republic has played a crucial role as one of the earliest publication centres of newspapers, «corontoes», «pamphlets», «mercuries», and other periodicals, intended for a local, national and even international audience. Dutch news publications are to be found in libraries and archives all over Europe, from London to Moscow. Newspapers in particular constitute an important historical source, because of both their contents and serial format. Until now, a bibliographical overview of Dutch newspapers has not been attempted, even though – as many as thirty years ago – the English historian G. C. Gibbs, in a seminal essay, characterized the compilation of such a catalogue as «certainly long overdue».

The Koninklijke Bibliotheek (National Library), The Hague, and the Persmuseum (Press Museum), Amsterdam, now have joined forces in formulating a project proposal for a digital bibliographical catalogue as well as the digitization of newspapers published in the Netherlands during the period 1618-1869. In 1618, the first known Dutch newspaper was published in Amsterdam, while in 1869; the special tax on newspapers was abolished, marking the end of the pre-industrial era. The envisioned catalogue will be a dynamic digital database, listing titles, copies, and holdings in Dutch and foreign libraries and archives. Since copies of many newspapers (especially from the early modern period) are not preserved in the Netherlands, and complete sets of individual titles are as a rule not to be found in one location, the project will also include microfilming and full text digitization. Thus, these scattered
sources will finally become accessible for future generations of historians, either online or in microfilm form at the Koninklijke Bibliotheek and the Persmuseum.

A pilot project, covering the seventeenth century, has been developed as a starting point. In order to compile a catalogue as possible, and to guarantee an optimal accessibility to the newspapers themselves, five non-consecutive steps have been marked:

- drawing up an inventory of files
- localization of holdings in Dutch and foreign libraries and archives
- microfilming and digitization of copies lacking in Dutch libraries and archives
- bilingual cataloguing of titles according to international bibliographical procedure
- providing access to full-text digitized Dutch newspapers.

This project is, per force, an international and an interdisciplinary enterprise. Therefore, the project team aims to establish a virtual community – an international platform for cooperation and (scholarly) discussion via the Internet. Its members will partake in the five steps cited above, as well as in research stemming from the project. This virtual community also offers exciting new possibilities with respect to educational projects, (online) publications and peer review. International conferences, to be organized during and at the end of the pilot project, will allow ample opportunity to discuss and assess the progress of the project, while at the same time providing new insights and theories in the field of news and media history.

Addresses for more information:

- National Library, The Hague, Mr. Jan Bos; jan.bos@kb.nl
- Press Museum, Amsterdam, Mrs. Angelie Sens; asens@persmuseum.nl
- or: Mr. Otto S. Lankhorst, Nijmegen; o.lankhorst@ubn.kun.nl

Information from a press release by the Koninklijke Bibliotheek and the Persmuseum

ICON (INTERNATIONAL COLLECTION ON NEWSPAPERS)

Report on the February 6, 2004 Advisory Board Meeting, Chicago, IL, Center for Research Libraries, by Missy Roser, ICON Project Coordinator

Update on Current ICON projects:

Preservation Microfilming:

Preservation is nearly completed for the approximately 21 titles. It is estimated that 255 reels will be produced. These include (some were previously reported in Newslettner no. 11): Argentinos lietuvi balsas 1949-77; Husaber 1952-58; Laikas 1950-76; Musu Postoge 1949-76; La Prensa 1955-60; Antorcha 1935-49 (in process); Caracas Journal 1946-49; Hispano americano 1943-48 (in process); Hombre Libre 1933-41 (in process); Omega 1919-37 (in process); El Popular, 1938-49; Sawt al-Jamiah 1972-76 (in process); Subhi aZadgan 1980-82; Tico Times 1973-80; Turkish Daily News 1962-80; Elore 1958-1971; Glas Slavonije 1946-1956; Ilustrirani vjesnik 1945-1952 (in preparation); Kulturni tvorba 1963-68; Lidova demokracije 1945-1962; Literaturuli Sak’art’velo 1963-1988; Magyar nemzet 1940-49, 1956-61 (in preparation); Odlej 1904-05, 1947-49, 1969-72 (in preparation); Slovenija 1932-1941; and Ujszo 1945-1955 (at filmer).

Survey of Newspaper Collections

The survey was developed in May 2003, with feedback from Advisory Committee. We sought to base it on the brief IFLA survey, but enhance it to obtain feedback on a variety of issues:
- Statistics of collection
- Collection and retention policies for international newspapers
- Storage and handling of newspapers
- Depth of cataloguing
- Reformattting and ILL policies

The final survey was sent to over 300 institutions and 2500 serials librarians, through direct mailings to CRL members and targeted lists of Carnegie-classified doctoral/research universities and major liberal-arts colleges; and through SERIALST, a discussion list for serials issues. 40 institutions responded, and summary results were made available at http://www.crl.edu/PDF/ICON_Survey_Results.PDF

The survey will remain available online for libraries to respond. Targets for additional responses include ICON members and major (suspended) holding libraries. Other suggestions made, in addition to those on the handout: Yale, U. British Columbia, Queens (Kingston), Toronto, Pennsylvania State.

A suggestion was made to translate the results of the survey. Further distribution of survey results should be done with suggestions for links to:
- NEWSPAPERS listserv
- IFLA Newspaper section list
- MICROLINK-L
- RLG-PRESERV

In terms of targeting future cooperation, a number of institutions suggested they would be interested in cooperating.

Cataloguing

252 original records were essentially subcontracted to University of Washington. 179 CONSER-level records have been completed so far (the bulk by Steve Early, with a first batch from Missy), approximately 80 more records have been identified to complete before the grant period ends, and 82 Custom foreign newspaper titles will need to be catalogued, for a
Continuing problems for cleanup include resolving issues with initial articles and nonfiling characters; ensuring successful searches on alternate titles and other search bugs, and work on diacritic search and display problems. New documentation has been produced for the SQL tables utilized and the MARC record to database mapping.

The batch-loading and editing interface is in development, with plans for a prototype to be ready by late spring. The process would involve running MARCedit and -breaker on each batch and doing any necessary cleanup and replacing before loading through a web-based interface. Ideally, the whole process would eventually be automated: each institution would run a saved OPAC report and search strategy before FTPing the file to the ICON server to be run against SQL queries. There exist possible open-source desktop PostgreSQL administration tools.

We are working out process for updates and input right now. We will need to institute regular updates once we figure out the process and discuss with LC CDS. We need to better promote the resource, with information on how to join and assist and search tips or tutorial. Also, information is needed about what is currently in the database (expectation of finding results).

ICON website

Because of the complexities of the existing management structure, ICON recently underwent a minor transformation of its Web site. The files created in Fireworks were essentially image files, leading to difficulties in changing information and enabling Web crawling. The new files are now text files, adhering to the same design of the current site. This was step one of an ongoing effort to improve the site and its usability. Now that the pages are searchable by Google and other engines, we will seek to release it more widely.

We also will be expanding the amount of information considerably, in line with our goals of providing a clearinghouse function by ICON. We have included a list of potential topics, and seek feedback. The need for feedback and additional reporting functions to monitor use is clear. We will examine the sorts of information and policies that should be added in terms of clearinghouse function and whether or not there can be at least one alternate language for the site and a way to call for gap fillers for filming projects.

Plans for next grant

Preservation microfilming 15 titles from Africa, Middle East, Latin America, and Eastern Europe have been identified for filming. Material was selected through a variety of tools, including assessment of ICON database information, a review of LC’s Foreign Newspapers list from 1954, and solicited participant and CRL member suggestions. 

Cataloguing

The cataloguing component of the next phase is based on a distributed approach, utilizing surrogates and student assistants to collate and note information about newspapers for bibliographic description and creation of exact holdings statements for 4 participating libraries. A second initiative involves distributed record identification and/or cataloguing for 10 institutions through the use of the ICON database. In describing these joint ventures, the question of what incentives exist for participation and continued support was also raised.

Database

As we move forward with adding new data loads, the need for a discussion of responsibilities for participants and better overall guidelines and documentation is evident. We’ll also be assessing interest in and recommendations for incorporating other databases (stand-alone, aggregated) as well as OAI harvesting.

Website

One of the major additions to the website will be an online microfilming queue. Among the issues to discuss are how much information to include, who should participate, and how? Should there be separate queues for current and retrospective newspaper filming? Another section of the «newspaper information clearinghouse» area will involve the conversion of finding aids, and we’re still researching format, content possibilities, and copyright issues.

Sustainability and long-term plans

Bernie Reilly reported on CRL’s PAPR (Preserving America’s Printed Resources) conference and followed with some discussion of ICON’s role in print retention and digital newspapers as well as selection issues for preservation projects. There is a possibility of using OCLC cataloguing credits for preservation services; what other self-supporting income exists?

The scale of the problem is so large, when one looks at the attrition of the world’s libraries, that we need to reconcile
our policies with the desire to retain as much as we can. ICON is an important program to the Center, as we focus more strongly on foreign resources. PAPR brought together print-based programs (storage facilities, conservators) to discuss how to manage the transition to digital resources.

We are looking at various funding models in the sustainability of ICON, including a member-funded model; a subscription model (providing additional services such as ILL, other delivery mechanisms); and a partnership model (with commercial entities, non-commercial groups, national libraries, etc.). We also need to consider copyright and fair use of the material in our distribution. Can we negotiate a better deal with the media community than we have now?

There is considerable interest in expanding this to be a truly international group. What are the incentives for foreign participants? Information sharing (also with IFLA), funds provided for preservation and holdings information, developing collections (featuring multicultural influences)?

Is there a possibility for an endowment? This would require a clear and compelling message, plus a strategy for activity. We have not fine-tuned ICON’s message precisely yet – something like «Preserving the biodiversity of knowledge»?

Reilly focused on the importance of the word «Coalition». We need to act decisively as a group, to better speak on behalf of our constituents and countries.

NEWS FROM THE IFLA SECTION ON NEWSPAPERS

HOT OFF THE PRESS:

New IFLA Publication: Newspapers in International Librarianship

This is a compilation of 35 papers given during the IFLA conferences during the 1990s. This is the period that the IFLA Round Table on Newspapers has been active in promoting awareness of newspaper collections in libraries, providing access to these collections and preserving them. An index of newspapers was added to facilitate access to the rich data in this book.

Newspapers in International Librarianship
Papers presented by the Newspapers Section at IFLA General Conferences.
(1FLA Publications; 107)
ISBN 3-598-21837-0
Price: EUR 78 / IFLA Members EUR 58
Order: K.G. Saur Verlag P.O. Box 701620 81316 Munich, Germany Tel: +49-89-76902-300 Fax:+49-89-76902-150/250 E-mail: info@saur.de http://www.saur.de

The editor of the News from the IFLA Section on Newspapers would like to invite you to submit any articles, announcements, upcoming conferences, and news in general on the topic of providing and preserving access to newspapers and the news in your country or region. Please contact Sandra Burrows at sandra.burrows@lac-bac.gc.ca or write to: Sandra Burrows, Newspaper Specialist, Library and Archives Canada, 395 Wellington Street, Ottawa, Canada K1A 0N4.

THE IFLA CONFERENCE 2005

will be held in Oslo, Norway. The Newspapers and Preservation and Conservation Sections are planning a joint meeting and tour of the National Library of Norway in Rana and its storage vaults on the 11th and 12th of August 2005. A tentative title for the program has been suggested:

Keeping Cool: Preservation Storage at the Arctic Circle.

NEWSPAPERS IN CENTRAL AND EASTERN EUROPE

The proceedings of the Berlin IFLA Post Conference Newspapers in Central and Eastern Europe (August 2003) have been prepared for printing; they will be published as an IFLA Publication in the IFLA Publications series in autumn 2004, thanks to the kind agreement of IFLA Headquarters and the publisher. This is remarkable as IFLA Publication 107 is dealing with newspapers already. On the other hand, it is a proof of the fact that the international interest in newspapers is growing and that the IFLA Newspaper Section is active and enterprising. The volume comprises almost 240 pages. The main focus is on issues of acquisition and access, copyright issues and digitisation and electronic newspapers.

Thus there will be another informative and attractive contribution to international newspaper work available soon!