International Preservation News

A Newsletter of the IFLA Core Activity on Preservation and Conservation

IFLA PAC

No. 56

May 2012

Digitization and Preservation of Newspaper Collections

INTERNATIONAL PRESERVATION NEWS
Nº 56
May 2012

ISSN 0890 - 4960
International Preservation News
is a publication of the International Federation of Library Associations and Institutions (IFLA) Core Activity on Preservation and Conservation (PAC) that reports on the preservation activities and events that support efforts to preserve materials in the world's libraries and archives.

IFLA-PAC
Bibliothèque nationale de France
Quai François-Mauriac
75706 Paris cedex 13
France

Director:
Christiane Baryla
Tel: ++ 33 (0) 1 53 79 59 70
Fax: ++ 33 (0) 1 53 79 59 80
E-mail: christiane.baryla@bnf.fr
Editor / Translator
Flore Izart
Tel: ++ 33 (0) 1 53 79 59 71
E-mail: flore.izart@bnf.fr
Spanish Translator: Solange Hernandez
Layout and printing: STIPA, Montreuil

PAC Newsletter is published free of charge three times a year. Orders, address changes and all other inquiries should be sent to the Regional Centre that covers your area.
See map on last page.

IPN is available on line at:
www.ifla.org/en/publications/32

IFLA-PAC Mailing List at:
http://infoserv.inist.fr/wwsympa.fcgi/info/pac-list

Any request for distribution should be addressed to:
flore.izart@bnf.fr

Front cover:
21-1-14, Chamonix : couple de skieurs, l'homme lisant un numéro du Journal. Agence Rol. 1914. Bibliothèque nationale de France, dpt Estampes et photographie, EST EI-13 (330).

ISO 9706

Digitization and Preservation of Newspaper Collections

Chronicles of Preservation

In April 2012 the Bibliothèque nationale de France hosted in Paris the annual IFLA Newspaper Section Conference entitled: *Newspaper Digitization and Preservation New prospects. Stakeholders, Practices, Users and Business Models*. More than 200 people attended the event which indicates that the theme was on everyone's mind.

Les marcheuses de la faim lisant un journal communiste.
Planet News. 1932. BnF, dpt Estampes et photographie, EI-13(2972).

Given that PAC was one of the organizers, one full day was reserved for preservation subjects (from paper conservation to digital preservation).

Proceedings of the conference will be published online[1] however we have decided to dedicate this IPN to several topics linked to the preservation and digitization of newspaper collections. Despite the fact that the articles contained in this issue are essentially looking at newspapers, the reflections and proposals outlined herein are applicable to other printed collections. My first meeting with the IFLA Newspaper Section standing committee (Salt Lake City, 2006) showed me a fundamental truth: given the complexity of digitizing press collections and the sheer number of pages involved, to succeed in this endeavour will be to solve most of the problems associated with the digitization of books. In this issue you will not find all the solutions but we would like to present an international selection of approaches concerning preservation and digitization: access and economics are touched on within this framework.

Two classic programs are presented: a large public library (Shanghai) and a major European publisher *(Corriere della Sera)*. The Library of Congress network is discussed by Deborah Thomas and Mark Sweeney. The general topic is introduced by Edmund King, former Head of Newspaper Collections at the British Library. I would like to take this opportunity to thank Ed for his invaluable contribution to the British Library Newspaper Collections and to the IFLA Newspaper Section, where he was Secretary.

I trust you will find the issue interesting and enjoyable.
IPN 57 (August 2012) will be dedicated to mass storage and preservation.

Christiane Baryla
IFLA-PAC Director

1. www.ifla.org/en/node/5932

L'Auto : les rotatives du journal sportif. Agence Rol.
1914. BnF, Département Estampes et photographie, EST EI-13 (351).

Du 11 au 13 avril 2012, la Bibliothèque nationale de France a accueilli à Paris la conférence internationale de la section Journaux de l'IFLA, dédiée cette année à la *Numérisation et à la Conservation des collections de Presse*. Plus de 200 personnes ont assisté à cet événement, preuve de l'intérêt suscité par ce thème.

Le programme PAC était l'un des organisateurs et une journée fut consacrée à la conservation.

Les communications et les powerpoint de la conférence sont publiés en ligne[1]. En parallèle, nous avons choisi de consacrer ce numéro d'*IPN* à la préservation et la numérisation des collections de journaux. Bien que les articles qui vous sont ici proposés traitent essentiellement de la presse, les analyses et les propositions présentées sont applicables à d'autres collections. Ce que m'avait appris, en 2006, la première réunion du comité permanent de la section Journaux de l'IFLA à laquelle j'ai assisté, c'est qu'étant donné la complexité et la spécificité de la numérisation de la presse et l'énorme volume de pages concerné, réussir dans cette entreprise reviendrait à régler la plupart des problèmes liés à la numérisation des livres.
Ce numéro ne propose pas un catalogue de toutes les solutions existantes mais une sélection d'approches internationales: les questions d'accès et celles des modèles économiques seront également abordées dans ce cadre.

Deux grands programmes classiques sont présentés: celui d'une grande bibliothèque publique (Shanghai) et aussi celui d'un groupe de presse européen majeur (le *Corriere della Sera*). Suit la présentation du programme de numérisation partagée à la Bibliothèque du Congrès par Deborah Thomas et Mark Sweeney.

Ce dossier est introduit par Edmund King, jusqu'à très récemment Directeur des collections de Journaux à la Bristish Library. Je souhaite profiter de cet éditorial pour remercier Ed de la contribution qu'il a apportée à la British Library d'abord et aussi à la section Journaux de l'IFLA, dont il fut le secrétaire enthousiaste et efficace.

J'espère que le thème de ce numéro vous intéressera. *IPN 57* reprendra en l'élargissant le thème de la session PAC à la conférence générale IFLA d'Helsinki : Magasins et Conservation.

Christiane Baryla
Directeur d'IFLA-PAC

1. www.ifla.org/en/node/5932

Old Wine in New Bottles: Historic Newspapers Online

by Edmund King, Former Head of Newspaper Collections, British Library, UK

Introduction

We have witnessed an extraordinary phenomenon in the last ten years. This is the incredibly rapid scanning of millions of printed documents and their availability via databases for users all over the world. The sheer number and variety of what is now offered is probably beyond the comprehension of any one individual. There is simply an incredible amount of information relating to older printed texts online. What is striking about this worldwide process of newspaper back run digitisation is the re-assertion of the need to present the whole text of newspapers, as originally published, to online users. So, we have old wine (the original newspapers), now placed into new bottles (the online environment). Of course, the big difference for users now is that many of the scanned newspaper texts have been converted via optical character recognition, so that text searching may be done by users. This has been stated by many others to be transformative for research, and I believe this still to be the case. Researchers now have the ability to search rapidly and purposefully to compare newspaper reports, opinion and evidence remotely on a computer. There is far less need for researchers to travel to the library of the archive that holds the original texts.

There is the need to respect the integrity of the original as originally created and published. We need to remind new audiences of how and why newspapers have been published over time; the social and historical context of their publication. And also continue to encourage the viewing of newspaper articles and snippets in context of the whole page or the whole issue, as originally published. The original entity is the issue. So this is the "old wine".

The transformation in researchers' ability to compare and to interpret newspaper texts, via online sources is itself a huge change. This can be done without too much difficulty online now, after some 10 years of worldwide effort in scanning and post-processing of old newspaper pages. I shall be offering a few case studies later in this paper. The difficulty now is dealing with digital excess, and working out precisely what we want to find from within the large quantities of results that emanate from a wide search, as part of a directed, thought out research event or item.

The examples being described next are mostly the means to an end, rather than the end of itself.

It is worth emphasising that the examples that follow are only drawn from sources or organisations that have enabled free public access to older newspaper texts. For this paper, this meant using newspapers made available by some national libraries: The Library of Congress[1], The National Library of Australia[2], the Bibliothèque nationale de France[3]. Google Newspapers archive is also available free[4]. There are considerable resources available from commercial companies also for historical newspaper research online, for example ProQuest Historical Newspapers[5]; Gale 19th Century newspapers[6]; Readex historical newspapers[7]; Newsarchive[8].

It is the original texts that are being presented and studying them in context remains as important as ever. The study of how and why information is presented on each printed page in the way that editors originally decided, is only just starting. Also, there is the possibility of comparison of newspapers with other original sources, of books, of prints, of maps, of photographs, of drawings, of paintings – all of this detailed work has barely begun. It is also a truism that some of the results could never have been secured by traditional reading of each newspaper page, looking for relevant information. Some of the results now resulting from online searches can only happen because of the whole scale indexing of older newspaper texts. The mass of texts available shows that information about events or people is only available in newspapers, often local ones in the country of origin. The means to search online and the relative ease with which researchers can accomplish this, are now realities.

A few of worked examples show how much the telegraph and cables linking countries, with consequent developments in communications, influenced reports printed in newspapers, especially from the 1850s onwards.

American Civil War – Loss of the Sultana

In the vastness of the American Civil War, occupying as it did three and a half years over a huge area of the United States, one incident will suffice to illustrate the potency of online searching. The loss of the steamboat, the *Sultana* in 1865 resonated across America. The SS *Sultana* was a Mississippi River steamboat paddle wheeler whose destruction in an explosion on April 27, 1865 was possibly the greatest maritime disaster in United States history. An estimated 1,800 of the *Sultana*'s 2,400 passengers were killed when three of the ship's four boilers exploded and the *Sultana* sank near Memphis, Tennessee.[9] Most of those on board were exchanged prisoners of war on their way home after privation and suffering from one to twenty-three months in Cahaba and Andersonville prisons.

1. *Chronicling America:* http://chroniclingamerica.loc.gov/
2. Trove. Digitised newspapers and more. http://trove.nla.gov.au/newspaper
3. Bibliothèque nationale de France: http://www.bnf.fr/fr/collections_et_services/presse/s.presse_numerisee.html?first_Art=non
4. Google newspapers: http://news.google.com/newspapers
5. Proquest Historical Newspapers: http://www.proquest.co.uk/en-UK/catalogs/databases/detail/pq-hist-news.shtml
6. Gale Cengage. 19th century US newspapers. http://mlr.com/DigitalCollections/products/usnewspapers/
7. Readex. America's historical newspapers: http://www.newsbank.com/readex/?content=96 ; African American newspapers: http://www.readex.com/readex/product.cfm?product=308
8. Newsarchive: http://www.newspaperarchive.com/defaultv11.aspx
9. Text from http://en.wikipedia.org/wiki/SS_Sultana

1. *Sheffield Independent*, Thursday May 11, 1865, page 2 – "the Sultana exploded on 27 April 1865".

The UK press picked up the story; no less than 73 articles appeared in UK London and Regional newspapers in the month of May 1865[10]. Of these, three appeared as early as the 11 May, when the *Sheffield Independent* reproduced a telegram sent to *The Times*. More details were available after another week, by the 18 May 1865. The account drawn from the *Memphis Bulletin* and printed in the *Fife Herald*, gives readers much more information about the nature of the incident, of the number of victims, and mention of how the rescue was conducted.

2. *The Nashville Daily Union*, April 29, 1865.

Looking at the *Chronicling America* titles, we have a brief headline in the *Nashville Daily Times*, of the 29 April 1865 which simply prints two headlines: "Steamer Sultana Burned. Great Loss of Life." (Such is the juxtaposition of different subjects within newspapers, we have this announcement in the midst of many commercial advertisements.) Newspapers in Australia picked up on the story as well. The NLA website has two newspapers which printed the story by July 1865. *Gippsland Times* (Victoria) carried the story in its issue of Sat 22 July 1865. *The Empire* (Sydney) printed the story on Wednesday 26 July 1865, p. 2. Perhaps one of the first analyses that a researcher can carry out is the differences between each of these reports; and also to identify the common source of the information, and what the original source printed in the first instance of reporting the story.

10. Figures taken from the *British Newspaper Archive*. http://www.britishnewspaperarchive.co.uk/ (Searched 21.12.2011)

The Siege of Paris during the Franco-Prussian War

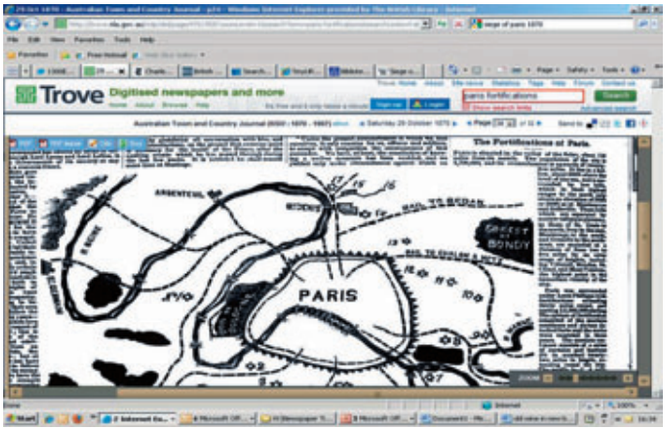
My second case explores the use of maps. Newspapers printed many maps during the 19th century. The Siege of Paris during the Franco-Prussian war attracted much newspaper coverage. As the Prussian armies moved into the interior of France, it became clear that Paris would need to defend itself. The Prussian army invested Paris between the 15-19 September 1870. *The Leavenworth Weekly Times* printed a map of the fortifications of Paris in its edition of 29 September 1870[11]. This is very detailed and is accompanied by a large article describing the layout in details. As we view it today, the centre looks overinked, but the roads of the periphery are well delineated. On Saturday 29 October 1870, page 24, the Au*stralian Town and Country Journal (NSW)* printed "The Fortifications of Paris"[12]. This is more of an overview of the area around the city, with the fortifications themselves shown only in outline. In the UK, the *Western Times* published this map of Paris and its fortifications on the 3 January 1871[13]. This is a carefully crafted map, with many locations named and the road system clearly outlined. Perhaps of particular interest to the researcher are the differences in presentation resulting from editorial choice at the time, this may have been linked to the limitations of what could actually be engraved for mass reproduction in each particular newspaper.

3. *The Leavenworth Weekly Times*, September 29, 1870.

11. *The Leavenworth Weekly Times*, September 29, 1870. http://chroniclingamerica.loc.gov/lccn/sn84027691/1870-09-29/ed-1/seq-1/;words=Paris+PARIS+FORTIFICATIONS+fortifications?date1=1870&rows=20&searchType=basic&state=&date2=1871&proxtext=paris+fortifications&y=18&x=9&dateFilterType=yearRange&index=2

12. Saturday October 29, 1870, page 24, *The* Australian Town and Country Journal (NSW) printed The Fortifications of Paris. http://trove.nla.gov.au/ndp/del/article/70462750?searchTerm=paris fortifications&searchLimits=l-decade=187llll-year=1870llll-monthInYear=October%7CmonthInYear%3A10 (Searched 21.12.2011)

13. Plan of Paris and its Fortifications. *The Western Times*. Tuesday January 3, 1871, page 7. http://www.britishnewspaperarchive.co.uk/ (Searched 21.12.2011)



4. The Fortifications of Paris. A*ustralian Town and Country Journal* (NSW: 1870 - 1907), Saturday October 29, 1870, p. 24.

Riot in Chicago in 1895

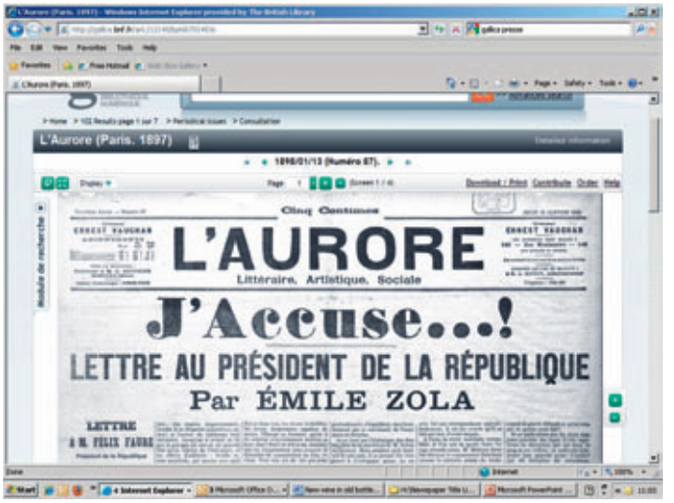
Another example is the closur e of the steel works r esulting from a demand for higher wages and consequent riot, that occurred in Chicago in 1895. *The Washington Times* reported the event on May 8, 1895; *the Standard* of London r eported the event on the 10 May 1895.

The Dreyfus Affair

In France, the Dr eyfus affair was an inter national event in the 1890s and early 1900s. Emile Zola' s letter to the Pr esident of the French Republic, a dramatic accusation, was the front page banner headline in *L'Aurore* of 13 January 1898 (no.87) [14]. On the same day, the UK regional daily, *the North-Eastern Gazette* ran the story with the headline: "M. Zola and the Dr efus Case." The fi rst sentence r eads: "M. Zola, in the pr omised letter on the Dreyfus Case, published in Paris today, says the affair is a stain and a blot on the Pr esident's term of offi ce." Only a day later, the 14 January, *The Huddersfield Daily Chronicle* runs the story with the bye-line: "Allegations against Fr ench generals".

5. *The Washington Times,* May 8, 1895.

14. *L'Aurore* front page, January 13, 1898 – Zola – J'Accuse. http://gallica.bnf.fr/ark:/12148/bpt6k701453s (Searched 21.12.2011)



6. *L'Aurore*, January 13, 1898 (no.87).

A search of the *Chronicling America* website under the term "Dreyfus Zola" (for all of the United States) yields 181 r esults for further scrutiny. This is a striking example of the kind of excess that researchers now face. Only thr ee days later after the publication of Zola's letter in *L'Aurore*, *The Washington Times* of the 16 January 1898 fers the headline: "Is All France Insane?" with the bye-line: "Unreasoning passion controls the Dreyfus Agitation". *The Kansas City Journal* of 18 January 1898 offers its readers a report of the events in Paris. *The Standard* of London the 19 January 1989 runs the story about how the Cabinet Council of France might take out a libel case against Zola; it also mentions the rioting that continues to take place. Australian newspapers reported extensively on the af fair, with a search of the words: "Zola Dreyfus" yielding 2,023 results for further scrutiny. A refinement of the search to article printed in January 1898 gives 118 results. The *South Australian Register,* Saturday 15 January 1898, carried the story on page 5. Interestingly, links are being made to other r esources available on the web as you can see from the left hand side of this screen shot. Commentary in Spanish newspapers was also extensive as we can see her e in *La Lectura Dominical* of 27 February 1898 [15].

7. Search of the *Chronicling America* website.

15. *La Lectura dominical* , 27/02/1898. http://hemerotecadigital.bne.es/datos1/numeros/internet/Madrid/Lectura%20dominical,%20La/1898/189802/18980227/18980227_00000.pdf?#search=%22zola%22 (Searched 21.12.2011)

8. *The Washington Times,* January 16, 1898.

9. *The South Australian Register,* Saturday January 15, 1898.

As we can glimpse, the coverage of this sort of event was very great, so early r efinement of online sear ches will yield fewer results, this enabling thinking more relevant to the needs of the researcher.

Conclusions

The work of deepening and intensifying the numbers of older newspapers online will continue. W ithin ten mor e years, we can expect to see much larger quantities of newspapers available online, from many more countries. There will be a mor e pressing need for our community of librarians and archivists to aggregate this information about all of them and to ensure that these listings remain available to all. At the same time, impovements to optical character r ecognition will be made, making researcher satisfaction all the greater. How to combine publicly available databases with those only available via charging will remain an area for further discussion. Her e, it has to be said that the price of entry to view commercially available databases will continue to lower over time, so this may not prove a barrier to use that it is perceived to be today.

The end of r esearch enquiries will be anything that the r e-searcher wants it to be, for example:

- Newspapers can be primary or secondary sour ces of information
- the pure enjoyment of finding fascinating information
- newspaper articles found which focus upon a particular historical subject, or person, or political movement
- articles found which show the number of reports about a particular incident
- articles found which of fer information and commentary upon social conditions (e.g. disease, or housing)
- articles found which permit the analysis of language itself in different countries, when describing the same event.

To deal with the pr oblem of sifting large masses of information, text mining techniques ar e being developed[16]. Semantic analysis for linguistics is now pr esent[17]. Both areas of activity will attract adherents, and many will spend much time focusing upon the means of attaining a summation of information using computers. The results of such work may or not may be successful, and may or may not lead to useful research within the arts and humanities. At this time, I pr efer to place trust in the diligence of the human mind and spirit (rather than in computer programs), in the ability of the mind to grasp the mass and complexity that results from massive amounts of information. Distillation of large quantities of textual information will be done by individuals or by teams of researchers, who combine knowledge of languages with knowledge of history , of literature, and of any other subject, to achieve r eadable summaries about a person, a subject or about events, fr om which we all benefit. This ability to synthesise and summarise will continue to benefit us all. The availability of vast amounts of texts (in this case for older newspapers) makes the journey more daunting, but hopefully more rewarding for those who undertake the challenge of research.

The last word on all of this – there isn't one!

16. Text Mining: http://en.wikipedia.org/wiki/Text_mining
17. Semantic analysis (linguistics): http://en.wikipedia.org/wiki/Semantic_analysis_(linguistics)

Digitization of *Corriere della Sera* Historical Archives

by Walter Colombo, DACS (Digitalizzazione Archivio *Corriere della Sera*) Project Manager, Italy

A Brief History of *Corriere della Sera*

The *Corriere della Sera* was first published in Milan, in 1876. The newspaper had four pages and was deliver ed in the afternoon. It had a cir culation of 15.000 copies. In 1883, a second afternoon edition was added; in 1890 a new mor ning edition was published, and finally, in 1903, *Corriere* became a four edition newspaper, two morning and two afternoon editions. The number of the pages was still four. In 1920 the circulation was raised up to 600.000 copies. In 1905 the pages doubled, and the eight page foliation changed again only in the fifties, when it got to sixteen. The number of pages then gr ew constantly, and was almost fi fty in 1992. Fr om 1867 to 1992 almost two millions of pages wer e printed. On 1945, April 25 th, the liberation day in Italy, *Corriere della Sera* stopped publication. After one month pages wer e printed again, but the name of the newspaper changed to *Corriere d'Informazione*. Finally, on 1946 may 7th, the name *Corriere della Sera* was printed again, and *Corriere d'Informazione* became the after noon edition(s) of *Corriere della Sera*, until the end of 1981. Since then only *Corriere della Sera* was printed, with all its editions. Today *Corriere* is deliver ed with a cir culation of about 480.000 copies, and has, besides local editions, fi fteen regional editions too.

The Documentation Center – Consultation and Preservation

In 1910 Luigi Albertini, Director of *Corriere della Sera*, decided that an internal library could be built to provide journalists of basic information tools. Casimiro Wronowski, a young copy editor, was chosen to build it. He started cr eating a collection of clipped articles, about people and facts. Since 1901 an index, year based, of the pages of *Corriere della Sera* had been created, containing subjects, names, places.

The *Corriere* Archive, due mostly to the annual index, has always had a double worthiness and destination: on one hand it was and is a consultation ar chive, with all the usual str ess coming along with the fr equent handling of volumes, on the other it obviously should be pr eserved, and so more than one copy of the volumes had been pr oduced. In 1974, a pr ocess started of creation of a microfilm archive, to become the main consultation archive

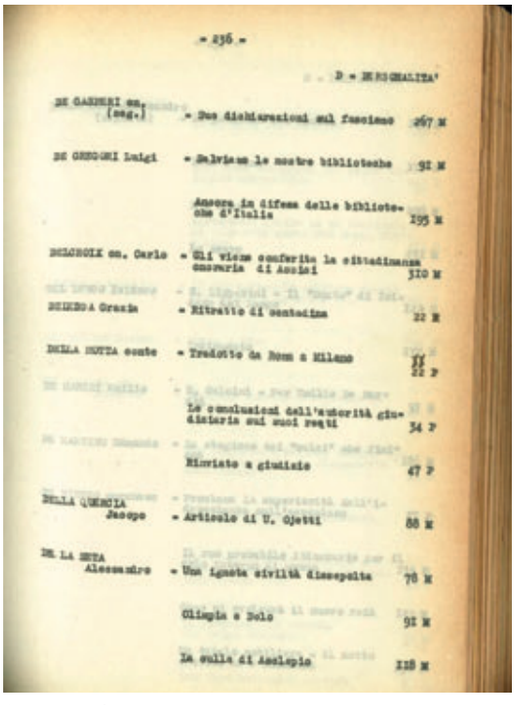
The Microfilm Archive

We must, in order to understand the way the microfilm archive was created, look at:

a) The edition structure of the pages
b) The need of internal use of the microfilm
c) The eventual external use of the microfilm

The *Corriere*, in the early years, as pr eviously mentioned, had just one edition, and got to four editions after 27 years. Unfortunately, of the first ten years we could not save all the different editions, and so all the year collections ar e bound in volumes that are not distinguished by edition, and contain all the editions we saved. Starting fr om 1888, volumes ar e identified by

1. The *Corriere della Sera* head office.



2. 1926 Index.

editions, where A and B are the first and second morning editions, and C and D (from 1903) are the first and second afternoon editions (photo 3). We must notice that, obviously, there are pages that do not change when the edition changes. The last page, for instance, that usually is an ad page, is the same in all the four editions. The cultural pages, too, and some others do not change very often. In defining the microfilm creation process, it was decided that:

1. The editions had to be the main organizational criteria.
2. Pages that did not change from one edition to another had to be kept only once.

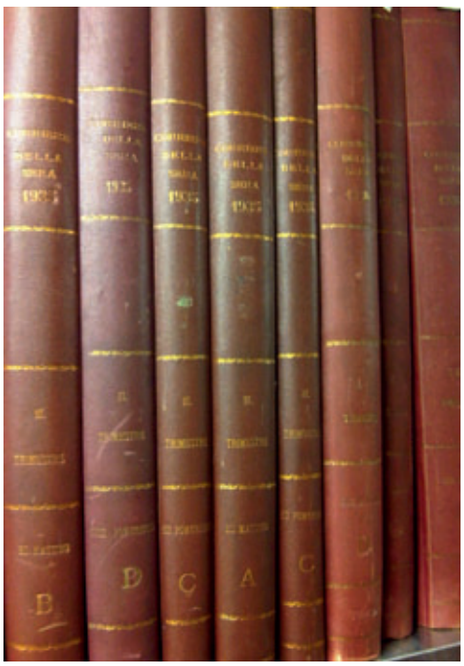
There was the need to choose the main edition, among the four, the one that had to be consulted in case of standard searches, when there was no need in choosing one specific edition, or when a microfilm had to be delivered to external organizations, as libraries.
Two other issues were defined:

3. Keep the two morning editions and the two afternoon editions totally parted, as if belonging to two different newspapers.
4. Treat the second morning edition (B) as the main edition.

This set, three families of microfilm were created:

- The first containing the second morning edition (B), with all the printed pages, and eventual supplement pages. This is called the B Microfilm.
- The second containing only the pages printed in the first morning edition and changed in the second morning edition. This is called the A Microfilm.
- The third containing all the pages of the fist afternoon edition (C) and the pages that in the D editions had been changed (in this case a specific page can be found twice, from C and D editions). This is called the C+D Microfilm.

The B microfilm is the main collection, and is used in internal standard searches and external deliveries; A and C+D Microfilms are used only for specific needs.



3. Volumes identified by editions: A and B for the first and second morning editions; C and D for the first and second afternoon editions.

This three family set does not change when, after 1945 April 25th, the *Corriere d'Informazione* takes the place of *Corriere della Sera*.

The Creation of Fondazione *Corriere della Sera*

In 2001, in order to save and give new value to the historical and cultural treasure of *Corriere*, above all the historical archives, the Fondazione del *Corriere della Sera* was created. This archive is maybe one of the most relevant archives in Italy, because of the wide time period that it covers, and quality and quantity of the several kinds of documents: correspondences between main personalities of political and cultural life in Italy, photographic archives, all the pages of *Corriere*. Moreover, following the ideal vocation of *Corriere*, the Foundation, always promoting meetings, debates and exhibitions, has become one of the main cultural issues in Milano, and in Italy too.

The Dacs Project

In 2009, promoted by Fondazione, the Dacs project started, with the help of ICT department of RCS Mediagroup. Dacs means Digitalizzazione Archivio del *Corriere della Sera*. The main objectives of the project are:

- Original paper pages conservation.
- Give more access to the pages.
- Increase the readership of *Corriere* while creating new business opportunities.

In the first phase of Dacs a panel with many vendors is established. The objective, besides the need of select a short list of vendors, was to better understand technical and logical issues of the project. In details, we wanted to know:

- If the scanning source had to be the microfilm or the original paper pages.
- The data structures needed to describe and manage all the objects of the project and their information. The importance of adopting some market standards.
- The maximum number of errors we could accept in the article texts, after the ocr process and post process, from a cost-benefit analysis point of view.
- The need of a complete workflow management engine in Dacs process management.

We decided that microfilm could be the better scanning source. You'll see below some details.
A Mets, Alto, Nitf xml structure, established market standards, was adopted to manage edition, page and article issues.
About text recognizing, we decided to establish a post ocr correction process, based on an integration between automatic and manual activities, the automatic correction process being managed by a semantic text engine.
We decided to adopt a complete workflow management system, to integrate and manage all the Dacs processes.
In 2010 a main contractor was chosen, and the Dacs project started its producing activities.

A central issue, in choosing microfilm as scanning source, was the decision of keeping, while structuring the digital archive, the same organization adopted for the microfilm archive. The

advantages of this decision are plain to see: the microfilm structure is already optimized, there are not two occurrences of the same page, because the selection task has already been accomplished. Furthermore, the microfilm scan, if managed with proper technology, is really faster and cheaper than paper scan. On the other hand, some microfilm pages may not fit the quality level we are expecting, because of some errors in microfilm creation. The average quality of the microfilm scan showed to be quite good as the paper (we have for all microfilm reels a master and some copies, and the master reels have never been used) but we had to be sure that we could track all the quality problems for all pages. So we put a control step, called *voting*, inside the workflow, in order to identify all the "below the minimum quality" pages.
The workflow steps, then, were:

Scan

The scan is performed on the entire microfilm reel, and the single pictures of each page are then made. The tiff files so obtained, one for each page, are then automatically processed by two crop and deskew tasks. These scanning and post scanning tasks, actually, are not really inside the workflow system, but the output tiff files are sent to the input queue of the workflow system, that starts handling them.

First ocr – page level – the voting process

As said before, we were afraid, and still are, now that the Dacs Project is in its second half, that some page scan quality was lower than we could accept. So we established a double step control for all the pages based on two criteria:

- Ocr process must recognize at least 90% of the text contained in the page.
- The semantic text engine dictionary must accept at least 80% of the words in the page text (of every kind of text, because there's no difference, at this point, between ads and articles and other things).

All the pages that do not go through this double examare refused, and an operator has the choice between rescan or ask for the paper. Until today we had about 0,5% of the pages refused, and we'd noticed that in most cases switching to the paper hasn't considerably improved the quality, being the paper itself the problem.

Page identification metadata insertion

Pages that go through the voting process are identified in term of date, page number, edition, etc., and some metadata are inserted in the database (if the page deals with sports, economics, or other predefined issues). Then a check is made about the completeness of the edition. Following the usual edition criteria, we decided that we needed a complete second edition, and that this had to be true for the morning and for the afternoon, and so:

- The morning editions A and B and the afternoon ones, C and D, are kept separated.
- We define the entities Last Edition (B and D) and Previous Edition (A and C). The Last Editions must have all the pages, and the Previous ones can be incomplete.

Page clipping – article identification: Last Editions only

Last Edition pages are clipped into text blocks, that are then gathered in articles, with each component defined, and the right reading order: header, title, subtitle, body text, author, pictures and captions. Obituary blocks are tagged, and only ad blocks are left without marking.

4. Manual text correction for Last Editions.

Second ocr – article level: Last Editions only

Last editions pages are then submitted to another final ocr process, that scans all the defined text blocks, in the right reading order, thus following the correct text flow in the articles

Automatic text correction by semantic dictionary: All Editions

Now we have on hand all the article texts of the Last Edition pages, and the page undistinguished text for the Previous Edition pages. The semantic text engine Cogito, by Expert System, analyzes all the texts and automatically corrects the wrong words for which the correction confidence is affordable and, for Last Edition articles only, tags as *wrong* the words that it considers may be wrong on semantic, syntactic or orthographic basis.

Manual text correction and final validation: Last Edition only

The last step in text correction is the manual one, and for Last Editions only. We use an application (managed by the workflow) that allows us to analyze and perform manual text correction in order to get to a 95% correctness of the body text and 100% correctness of the title.

Article linking: Last Editions only

This last step allows us to manually link all the articles that flow on more than one page, and this leads to put (see next paragraph) the complete text of this kind of articles in a single xml file.

Dacs deliverables (digital objects) and xml structure for the exhaustive description of the digital archive
The objects produced for a specific edition are:

Edition level (Last and Previous Edition):

- One Mets xml file. It contains all the edition metadata, and a description and reference of all the objects.

Page level (Last and Previous Edition). For each page:

- High resolution tiff file
- Jpeg file at several resolutions
- Pdf file with searchable text
- Alto xml file, mapping all the contained words and their tagging

Article level (Last Edition only):

- Jpeg file of the clipped article at several resolutions
- Nitf xml file containing the whole text of the article with its tags

The xml structure of Mets, Alto and Nitf are the standard ones, with some elements added: cross references between the three different xmls, and some editorial and semantic information in the Alto file, so that the Nitf file could be completely constructed starting from the Mets and Alto files.

Sustainability in the United States National Digital Newspaper Program

by Deborah Thomas and **Mark Sweeney,** Library of Congress, USA

Abstract

This paper describes the program organization, technical specifications, and tools that support the U.S. National Digital Newspaper Program (NDNP), a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC). NDNP is a long-term effort to provide permanent access to a national digital collection of newspaper bibliographic information and selected historic newspapers, digitized by NEH awardees in all U.S. states and territories. The program provides the Library of Congress with a testing ground for the development of large-scale distributed, digitization programs and for predicting long-term needs for management and preser vation of digital assets. The development focuses on creating digitized newspaper page surrogates through a distributed effort, ingesting the resulting digital objects into a system, providing user-friendly access to the data, while implementing a system that is capable of sustaining the content for future use.

Whether describing political rallies, disasters or r eal estate sales, historic newspapers are the primary record of events that chronicle the development of communities[1]. The U.S. National Digital Newspaper Program (NDNP), a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC), is a long-term ef fort to pr ovide permanent access to a national digital r esource of selected historic newspapers digitized by NEH-funded institutions from all U.S. states and territories and newspaper bibliographic information about collections across the country. This program builds on the legacy of the United States Newspaper Program (USNP, 1982-2011) sponsored by NEH with LC technical support – an excellent example of successful collaboration at both the national and state levels to inventory , catalog, and selectively pr eserve on microfilm a corpus of at-risk newspaper materials. The newer NDNP not only extends the usefulness of USNP bibliographic and microfilm assets by increasing access to this valuable information, but also provides an opportunity for many institutions to contribute select digitized newspapers to a fr eely accessible national newspaper resource.

Newspapers throughout American history have provided a venue for sharing the facts and opinions of moments in time, significant people, and local perspectives—a unique r esource for recording and understanding both singular and united voices on ideas, events, and democratic identity . In r ecent decades, under USNP, the preservation of newspapers on microfilm and the establishment of imaging and bibliographic standar ds has been an important component of archival programs, enabling

1. Newspaper pages fr om *Chronicling America: Historic American Newspapers*.

them to manage and sustain the vast quantity of material r epresenting the historic record effectively. However, even this critical aspect of newspaper librarianship does little to address the use and access needs of text-intensive newsprint. Even with the best imaging standards and process, the intellectual content of the newspaper is presented in a complicated layout, with varying visual cues and small typefaces. The development of new digitization technologies, text recognition, and search engines enables the NDNP to provide enhanced access and discovery to this material, as well as the national leadership necessary to establish best practices and standar ds for the digitization and structure for historic newspaper materials intended for a sustainable electronic resource.

Establishing the Program

The primary goals of the pr ogram are long-term – to pr ovide enhanced access to select newspapers by cr eating and aggregating millions of digitized pages fr om geographically-diverse communities while also repurposing existing bibliographic and

1. This paper is a revision and update of Mark Sweeney, "The United States National Digital Newspaper Program (NDNP): a distributed national ef fort to enhance access to America's historic newspapers." International Federation of Library Associations W orld Conference and Assembly, Milan, Italy, August 2009. Conference presentation.

holdings data for over 140,000 U.S. titles in a freely-accessible and searchable system. Since the U.S. national newspaper collection is dispersed among hundreds of libraries throughout the country, a decentralized selection and digital conversion model was adopted with data aggregation provided by the Library of Congress for access and preservation.

In 2004, NEH and LC began a collaboration to develop a nationwide program that enhanced access to this material through the use of new technologies and information channels, scaled to include representative content from all U.S. states and territories produced over several decades, and encouraged interoperability between digital libraries through shared specifications. A memorandum of understanding between NEH and LC clearly delineates the responsibilities of the two agencies in developing the overall national program. While NEH manages and funds annual award competitions among state-level institutions to select and convert historic newspapers to digital form, LC focuses on the program's technical specifications, data management and publicly serving the content. State level institutions, known in the program as "awardees," are responsible for selecting newspapers published in their state according to program guidelines and converting them to valid digital form for central aggregation at LC.

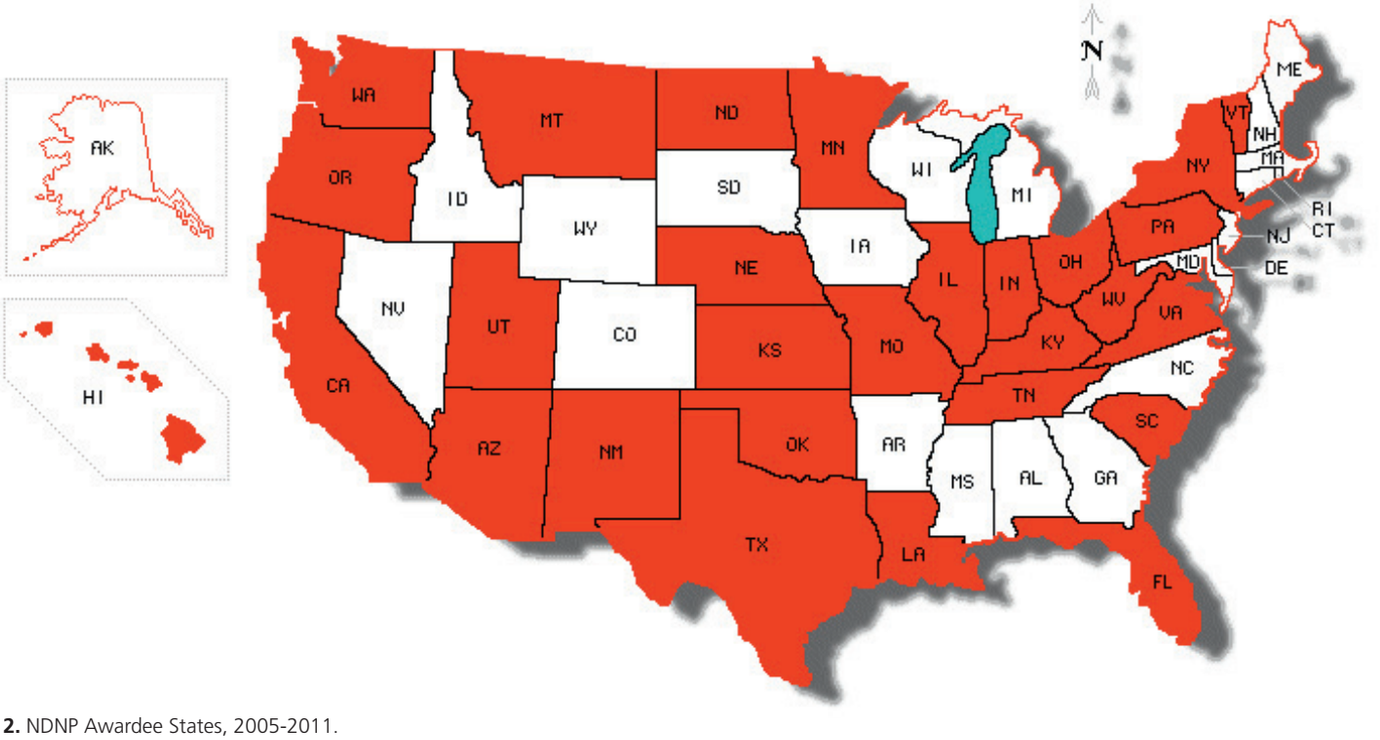
In 2005, NEH held a national competition for institutions to select, digitize and contribute 100,000 pages each from their historic newspaper collections to the NDNP program. In this first year, NEH selected six contributors based on their experience with historic newspapers, digitizing collections, and digital library infrastructures. In subsequent years, from 2007-present, NEH has made additional awards and supplemental awards to a total of 28 states, funding digitization of more than 5.6 million pages to date. A competition is currently underway for this year's awards to be issued in August 2012.

Technical Specifications and Tools

In the development and overall management of the program, the Library of Congress provides technical support of the program's primary goal – creating open access to the nation's historic newspapers. The Library's role has three parts: to establish technical digitization specifications that permit aggregation, to serve and unify this content through a publicly-available Web site, and to permanently sustain the aggregated content. As LC reviewed the means available to accomplish these objectives, it became clear that the requirements of sustaining the content over time would inform many decisions supporting the other two objectives.

The NDNP data management environment is based on requirements supporting the four major workflows as identified in the Open Archival Information Systems (OAIS) Reference Model: ingest, archiving, dissemination and preservation management. From the outset, LC recognized the scope of the planned program – millions of newspaper pages produced by many different organizations over several decades (equaling, at least, hundreds of terabytes) – and the commitment between publicly-funded agencies to manage access to this content required emphasis on the creation of digital assets according to emerging standards and uniform best practices. Well-formed data operating in a robust technical infrastructure would be the best approach to ensuring cost-effective management of the content over time.

The Library began with identifying high-level operating principles and functional requirements for a digital asset management environment that supported ingesting and archiving, as well as the associated dissemination workflow. In a climate of emerging (and evolving) best practices for digital preservation, LC initiated an explicit development phase to allow for research



2. NDNP Awardee States, 2005-2011.

and assessment of long-term workflow and curation needs, as well as incremental progress toward NDNP goals. The principles applied in making technical choices were intended to support the development of a system that is sustainable in today's best estimation – open, modular, certain to change, and able to evolve to meet future uses.

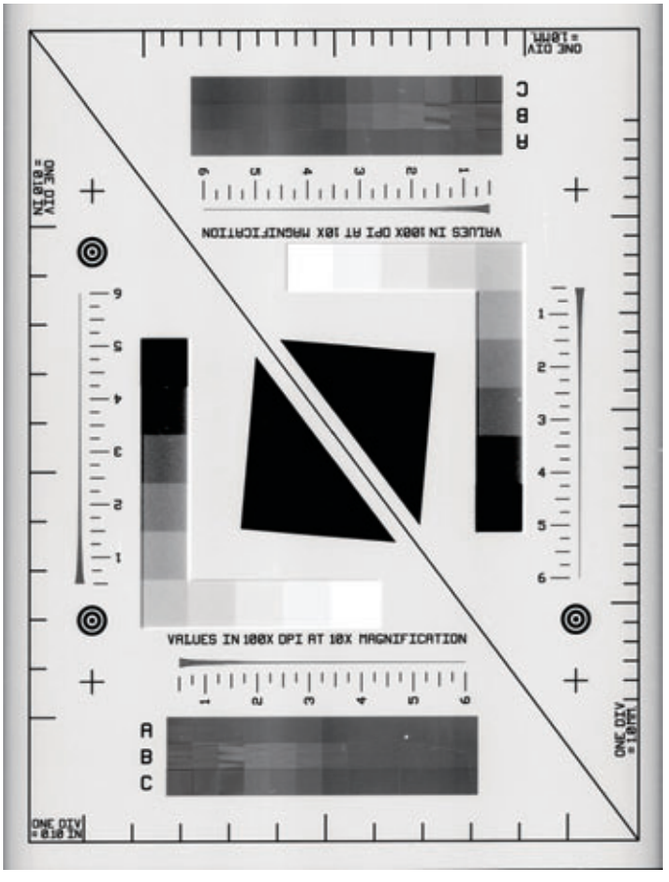
These technical decisions were also informed by realities of the overall program structure:
– The content – analog versions (microfilm, paper) of U.S. historic newspapers – resides primarily in state repositories, rather than at the national library, and therefore the program requires distributed production of the digital assets;
– The funding to apply new technologies to enhance access to this material is finite, therefore,
  • content included in the program will be selective, rather than the entire corpus available;
  • technical requirements for converted materials should account for potential re-use and reprocessing over time (scan once, use many times);
  • the program should provide a model for other efforts that may eventually interoperate – sharing best practices, conversion specifications, and standardizing basic access for historic newspapers.
– Demonstration of good use of public funds by providing open and perpetual access;
– In expectation of change, avoid closing off options, by developing a system environment that would be open, expandable, and modular.

To build an extendable, scalable, and sustainable workflow, eventually encompassing more than 50 producers, NDNP needed to consider five main requirements:
– convert the content to achieve the highest quality information for discovery and re-use,
– ensure technical consistency across multiple producers over time,
– use open and sustainable formats to encourage long-term preservation,
– develop a data architecture that would allow for both manageability and scalability over time, and
– develop scalable workflows, quality processes, and data transfer tools that support inventory and management of a large-scale digital collection, created by multiple producers.

Building on its experience with large-scale digitization of historic materials, LC developed technical specifications for NDNP content based on existing best practices. The image specifications – TIFF, JPEG2000, and PDF – are intended to play specific roles in the NDNP system (TIFF for archiving, JPEG2000 for production and PDF for portability) and conform to current best practices for digital file format sustainability[2]. These practices include wide-ranging adoption in the cultural heritage community, transparency of the digital information itself, and self-documentation within the file format. The image specifications for NDNP, primarily 8-bit grayscale at 300-400 dpi, attempt to capture the most data possible from newspaper microfilm negatives to provide for future reprocessing and reuse with improved technology. In addition, LC chose a standard XML metadata scheme (Metadata Encoding and Transmission Standard[3]) for description of the digital objects at the newspaper issue and page level and the ALTO (Analyzed Layout and Text Object) schema extension[4] for structuring the automatically-recognized machine readable page text (optical character recognition or OCR). Metadata requirements were intended to provide a basic level of access to newspaper pages, capturing as much structural and technical information as possible from both film and intellectual content at the point of digital creation.

NDNP recognized that a distributed production model would require improved mechanisms for quality assurance of the content as it was created and aggregated, as well as explicit incorporation of metadata intended to assist in long-term management and sustainability of the digital objects. These requirements led to the development of two NDNP-related tools, a microfilm scanner target for objective image quality analysis and a technical validation and quality review software, both used by program participants to assist in capturing technically valid, high-quality images and ensuring that metadata conforms to NDNP technical specifications.

The NDNP image specifications attempt to capture the most data possible from newspaper microfilm and the program has



3. Preservation Microfilm Scanner Target (PMT), image provided by Image Science Associates.

2. "Sustainability of Digital Formats – Planning for Library of Congress Collections." http://www.digitalpreservation.gov/formats/ (Accessed 22 March 2012)

3. Metadata Encoding and Transmission Standard (METS): http://www.loc.gov/standards/mets/ (Accessed 22 March 2012)

4. Analyzed Layout and Text Object Schema (ALTO): http://www.loc.gov/standards/alto/ (Accessed 22 March 2012)

established technical specifications and workflow components to that end. Microfilm is assumed to play the leading role in content selection since most original paper issues from the target time period have significant deterioration or are simply no longer available. The capture of a standardized image target along with the selected historic material is a best practice used by many digital library projects to further the goals of producing accurate materials that can be managed in the absence of the original item. Recognizing that no such test target existed for the digitization of microfilm at the time, NDNP worked with Image Science Associates[5] to develop the Preservation Microfilm Scanner Target (PMT), a standardized technical test target on microfilm (see Fig. 3) with associated analysis software, to assist in creating the high quality digital images that the program requires.

The PMT can be used to create a benchmark for scanning equipment capabilities or to support ongoing quality control during production workflow. An initial set of scanned target images from a specific capture device can establish a benchmark for anticipated performance of that particular device or variables associated with settings. Comparison of benchmark scans from different scanning equipment or different vendors can assist in making choices among them. In addition, if there are quality concerns for a particular image set (blur, contrast, noise, etc.) PMT images captured at the time of digitization can be useful in determining which variable is responsible for quality concerns and how to address them.

The film target and analysis software contains several elements to ensure consistency with current International Organization for Standardization (ISO) imaging specifications. In ISO 14524, the responsiveness of the capture device to tones is defined as its Opto-Electronic Conversion Function (OECF). The PMT contains a series of gray boxes, with graduated levels of darkness that should be distinctly observable in a high quality scanned image. This creates both a visual clue that the target image captured the full range of tones available, and data points that may be analyzed by software to calculate the OECF for the system. Given the variety of content elements in newspaper images, it is important not only that enough pixels per inch are captured, but also that the optical system captures enough detail to justify those pixels. In ISO 16067-2, sharpness is measured in the system's Spatial Frequency Response (SFR). The PMT contains a slant-edge border between white and black areas and eye-readable resolution charts consisting of narrowly spaced lines. This is the input for a software analysis of the system's SFR capability.

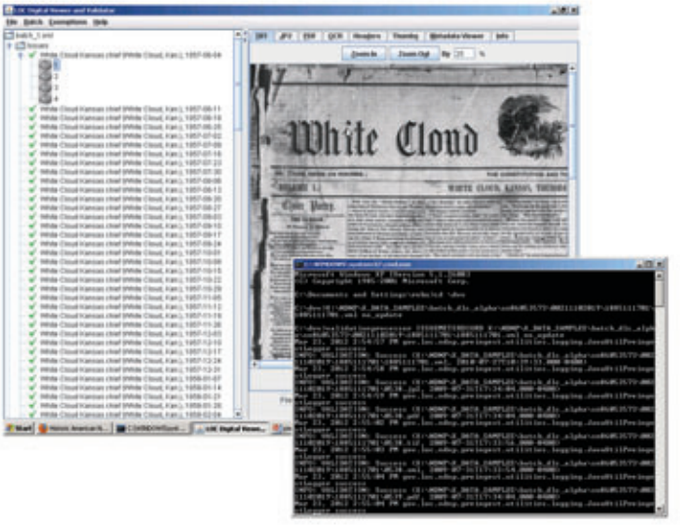
The process of digitization often produces unintentional noise: artifacts that appear randomly or systematically in the digital image that were not in the original. The PMT measures noise through the use of a series of squares with very small vertical and horizontal lines and a long diagonal line across the entire target. If there is interference created by the distance between pixels on the sensor and these lines, the target image will produce a pattern of widely spaced lines. The degree of fluctuation is indicated by the analysis software, which also provides information on the likely range of acceptable fluctuation.

5. See Image Science Associates: http://www.imagescienceassociates.com/, for more information (Accessed 22 March 2012)

For NDNP production workflows, the analysis of PMT target images may be done as often or as little as awardee project managers deem necessary, frequently for quality monitoring or on an as-needed basis. To date, analysis of scanned targets within NDNP has revealed issues in scanning performance in the areas of tonal reproduction, sharpness and noise, enhancing producers' ability to monitor aspects of the digitization process. Just as importantly, when poor newspaper images on the original microfilm caused the scanning process to be questioned, the use of the target has also demonstrated when the scanning devices is operating correctly.

In order to ensure conformance with other NDNP technical specifications, a second tool was needed for NDNP staff, awardee institutions, and digitization vendors alike to validate conformance to NDNP technical specifications and enable quality assurance processing. To support efficiency and scalability, technical conformance (e.g. whether a field is populated with the appropriate data type) would be supported by automated analysis while a data object viewer would support more subjective human inspection (e.g., whether the field data is correct).

To support automated technical conformance, much work had already been done at Harvard University with the creation of the open source JSTOR/Harvard Object Validation Environment (JHOVE) software[6]. This software was able to measure and characterize many aspects of the file types (JPEG 2000, PDF, TIFF) that are used in NDNP. Further programming at LC extended the capabilities of the software and incorporated validation of the XML metadata necessary for the NDNP system[7]. This analytical code was wrapped in a graphical user interface that became known as the Digital Viewer and Validator (DVV). The DVV then allowed automated analysis of the objective criteria



4. NDNP Digital Viewer and Validator, quality review interface and DOS command-line module.

6. JSTOR/Harvard Object Validation Environment (JHOVE): http://hul.harvard.edu/jhove/ (Accessed 22 March 2012)

7. For more explanation of the digital object validation strategies implemented for NDNP, see Littman, Justin. "A Technical Approach and Distributed Model for Validation of Digital Objects." In *D-Lib Magazine*, 12:5 (May 2006). http://www.dlib.org/dlib/may06/littman/05littman.html (Accessed 22 March 2012)

of the data objects ensuring the right data types were used, and a visual quality review to ensure the metadata was being employed correctly (e.g. the right title was used, the date in the metadata matches the date on the page image).

During validation, initiated through either the graphical interface or a DOS command-line module, the DVV verifies approximately 100 characteristics of the data package. If the files meet the specifications, it extracts header data from the various self-documenting file types for transformation into Preservation Metadata Implementation Strategies (PREMIS) and Metadata for Images in XML Schema (MIX) schemas within the associated METS object. In addition, the DVV adds a digital signature to the METS object for each associated file. This digital signature can be checked later, to determine if the file has changed in the interim, whether intentionally, by operator error, or by bit degradation. These features of the DVV allow the validity of NDNP data files to be monitored throughout their lifecycle.

Early in the program NDNP determined that to maintain a sustainable cost-effective program, the data produced by NDNP must be managed in such a way as to ensure both its reliability and quality for use. One early lesson was identifying the need to minimize human-interaction (and therefore human error) in the data lifecycle. To that end, when LC receives the NDNP data, valid and with digital signatures for each file intact, the data delivery is then "bagged" using the Internet Engineering Task Force (IETF) draft specification BagIt[8] (developed under the National Digital Information Infrastructure and Preservation Program[9]) and transferred into an automated workflow supported by generalized transfer utilities that provide management services to enhance reliability, findability, and integration with other digital collection material.

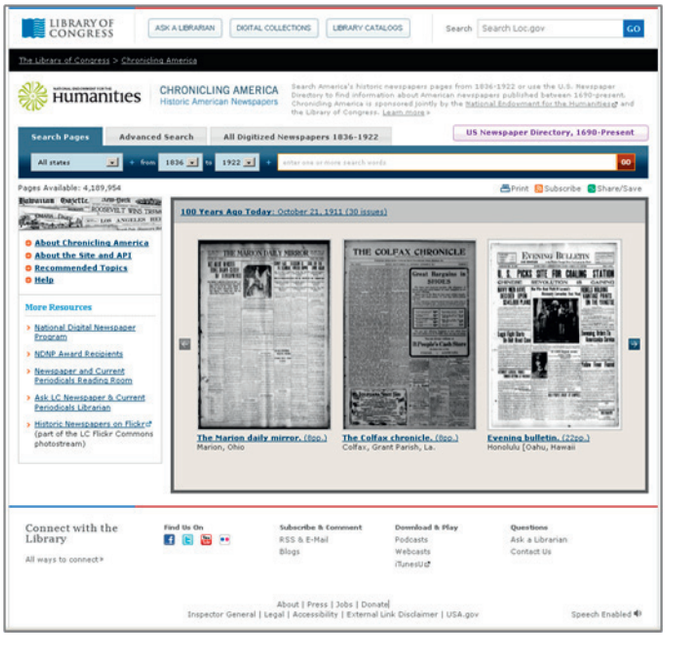
The tools described above help ensure the initial NDNP investment in data specifications and quality monitoring results in a sustainable resource worth the cost of management and maintenance over time.

Enhancing Access through Use: *Chronicling America*

During the early stages of the program, LC developed requirements for an access system through identifying use cases and scenario planning. The NDNP team defined needs for access as the ability of a general user to search and/or alphabetically browse newspaper title directory records, browse through various digitized titles by issue date and logical page order, and to support simple keyword search at the newspaper page level. Automatically recognized machine readable text (optical character recognition, or OCR) with associated word coordinate data, structured using the ALTO schema provided the basic page structure and keyword searchability as well as locational information that could be used in a visual interface to highlight search results. Additional structural or descriptive metadata identifying parts of the page was not included in the specification in order to maximize the available resources and meet the needs of providing basic access to the content. Formal usability testing was conducted on an early prototype incorporating this functionality to confirm general usage assumptions and needs.

The user interface to NDNP data, known as *Chronicling America: Historic American Newspapers* (http://chroniclingamerica.loc.gov), is freely accessible to the public and available from the Library of Congress Web site. The current Web site architecture supports human, machine, and API (Application Programming Interface) usage. Building on open source software toolkits, the architecture leverages the Apache HTTPD Web Server, the Django web publishing framework[10] (originating from the newspaper publishing sector and able to fulfill many of the digital resource use cases with very little additional effort), the JQuery JavaScript Library, MySQL database, Solr search server, and a number of Python libraries. These open source tools provide the flexibility, reliability and ease of use needed both internally at LC for implementation and for external users. For use outside the Library of Congress, the complete site application is packaged together for open-source distribution through SourceForge.net as "the LC Newspaper Viewer."[11] Most importantly, LC has substantively revised the site 2 times to date, enhancing access to the site and use of the site by multiple factors, without any changes to the NDNP digital object technical specifications or structures.

This architecture supports robust performance that allows web crawlers and search engines access to *Chronicling America* content. Providing machine access to the content has greatly enhanced use of the newspaper collection by putting it in front of millions from whom it was previously hidden. Implementing standard APIs allows external dissemination and curation



5. Screen shot, chroniclingamerica.loc.gov.

8. BagIt specification: http://www.digitalpreservation.gov/partners/resources/tools/index.html#b (Accessed 22 March 2012)

9. National Digital Information Infrastructure and Preservation Program: http://www.digitalpreservation.gov/# (Accessed 22 March 2012)

10. Django Web Framework: http://www.djangoproject.com/ (Accessed 22 March 2012)

11. SourceForge.net/Library of Congress – Newspaper Viewer: http://sourceforge.net/projects/loc-ndnp/?source=directory (Accessed 22 March 2012)

of the site content. For example, the OpenSearch API implemented allows users to search the newspaper pages and titles directly from their Web browser or harvest newspaper images, text or bibliographic data to provide insight into the historic progression of newspaper publishing in the U.S. Supporting the semantic web and RDF (Resource Description Framework) and using common linked data vocabularies, each newspaper page and bibliographic directory record has a persistent URL and includes Dublin Core and MODS data embedded in the page view as HTML. Presenting the NDNP data in multiple ways beyond the Web-to-user interface promotes use of the content via new research techniques such as data mining, visualization or machine-user methods of discovery.

The site currently includes approximately 5 million newspapers pages from more than 700 titles, published between 1836 and 1922 from 25 states and the District of Columbia, with 3 additional states pending. Also included is a directory of newspapers published in the US from 1690 to the present and information on libraries that hold them in both physical and digital form. Since March 2007, the site has provided content to approximately 7 million visitors with over 65 million page views. In addition to basic keyword search functionality, the Web site provides access to citation information for each newspaper page, visual calendars indicating available issues for a given year, files for download and re-use, special features for clipping (or printing) detailed images of a page, persistent "bookmarkable" links for all site views, newspaper histories for each digitized title, regular RSS (Really Simple Syndication) feeds of content highlights, updates, program developments, and more. Additional links are provided to share any Web page in the site by email and social media, such as Facebook, Twitter and others. Likely areas for future enhancement to access are language-specific search capabilities for multilingual page content (currently English, French, German, Italian, and Spanish are anticipated), additional automated analysis and manipulation of OCR to enhance search specificity and the overall user experience.

Sustaining the Content

An important component in the fulfillment of LC's role in this program is the development of a system environment that ensures the digital assets acquired from many different sources over a long period of time will be sustainable. The environment must guarantee that when people, process, and technologies change, the digital asset can be available (transparently and automatically if possible) for use and, potentially, transformation. Appropriate repository architecture is an essential component in determining if a digital preservation environment is successful.

As mentioned earlier, the core custodial unit of digital content at the Library of Congress is compliant with the BagIt specification where each unit is called a *bag*. The content itself is stored in files in the bag, which are listed with their cryptographic hash values[12] in a bag manifest. Bags also contain administrative metadata about content that help curators now and in the future reconstruct files into usable and useful information.

To provide repository services that ensure digital assets are preserved over time[13], the Library developed a suite of software and services referred to as Content Transfer Services (CTS). CTS moves bags through an automated workflow that ensures not only the safekeeping of the bag, but also provides checkpoints for subjective review through the end of the process. Upon receipt of deliveries, CTS performs a malware scan. To ensure the continued integrity of the content, no copies to any storage location are considered complete until CTS verifies the cryptographic hash values. Periodic spot-checks on content to detect accidental changes or corruption are also initiated using CTS. The workflow completes with a bag being copied to access servers, archival systems, and an automated process is kicked off to make the former available to the public.

Each operation in CTS creates and stores preservation metadata, which is accessible to curators through the CTS web interface. Users of the system can manually add or modify administrative and bibliographic information as content is being stored or manipulated. Staff can use the interface to browse directly to files without needing direct server access, but in a way that still provides an access control layer.

In addition to workflow and preservation management, CTS includes extensive reporting about both the content and its use which library staff can view through standard reports and ad hoc queries. Locations of data, storage used, file formats, files stored, and progress through automated workflow are all visible to curatorial staff. These reports are exportable to spreadsheets which are used in reports to program stakeholders, the U.S. Congress and to inform billing and contracts.

Library staff constantly monitor the risks[14] to the assets under LC stewardship and make regular improvements to CTS and its processes. Evaluation of the lessons learned, as well as input from other LC collection management projects, will continue to inform the development of additional tools and services for managing the complex workflows associated with data acquisition, management, access and archiving.

Supporting Infrastructure for Sustainability

The NEH and LC have made a long-term commitment to the development of this program and its digital assets, including a formal agreement regarding goals of the program, cost-sharing for technical development and data management, and cooperatively guiding the program's progress. In order to fulfill its role in providing permanent access to this high-value historic content, LC implemented supporting infrastructure – both pro-

12. For more information about cryptographic hashes and content integrity, see Zwaard, Kate. "Hashing Out Digital Trust". In *The Signal*, (November 2011). http://blogs.loc.gov/digitalpreservation/2011/11/hashing-out-digital-trust/ (Accessed 27 March 2012)

13. See Littman, Justin. "A Set of Transfer-Related Services." In *D-Lib Magazine*, 15:1/2 (January/February 2009). http://www.dlib.org/dlib/january09/littman/01littman.html (Accessed 27 March 2012)

14. See Littman, Justin. "Actualized Preservation Threats: Practical Lessons from *Chronicling America*." In *D-Lib Magazine*, 13:7/8 (July/August 2007). http://www.dlib.org/dlib/july07/littman/07littman.html (Accessed 27 March 2012)

grammatic and technical – to enable the long-term sustainability of the collection.

The organizational infrastructure established at LC included an internal program management team, made up of stakeholders representing collections interests, digital production (conversion and acquisition), and digital preservation. These stakeholders had hands-on experience in a broad range of LC programs, including newspaper collection development, the American Memory digital historic collections, Ameritech-funded partnerships, information technology and the National Digital Information Infrastructure Preservation Program (NDIIPP). Together, these committee members represented various management groups in the Library and successfully scoped the LC roles and deliverables that fulfill program requirements –administering a successful distributed production model, providing a Web interface to acquired data, and developing a system environment to maintain and sustain the digital content.

To accomplish these goals it was essential that LC also establish a technical development team representing various specialties - preservation architecture and repository development, data modeling, software development, search analysis and user interface development - willing to experiment and contribute to the advancement of best practices in digital preservation. This team shares expertise with other LC digital collection management efforts, including copyright deposit of electronic journals, Web archiving, and digitized historic collections, using and generalizing the lessons learned in NDNP to extend the repository services to other content types. The technical development group supporting NDNP supports not only the creation of a system environment that meets NDNP goals, but also the establishment of a repository development center (hardware, software, and systems) for on-going research at LC into the challenges of preserving all types of digital information.

Together, LC, NEH and the participating awardees have developed a strong community of practice around the digitization of historic newspapers that:

- allows for a uniform body of content created by multiple producers,
- meets basic user access needs beyond analog versions,
- has a high-degree of sustainability, and
- enhances access to American newspapers overall.

The immediate up-front decisions on the best practices and strategies for sustainability that would lead to a successful program have been validated. As the program continues to develop and expand, LC will continue to adapt and evolve the tools and systems supporting this program. Facing the challenges of building a national digital newspaper collection has informed universal understanding of needs and capabilities for the preservation of all digital information.

Acknowledgement

Thanks to David Brunton, Robin Butterhof, Liz Madden, Nathan Yarasavage and Kate Zwaard at the Library of Congress for contributing to this article.

Sostenibilidad del Programa Nacional de Prensa Digital de los Estados Unidos

Este documento describe la organización, especificaciones técnicas y herramientas del programa que patrocina el Programa Nacional de Prensa Digital de los Estados Unidos (NDNP, por sus siglas en inglés), una asociación entre el National Endowment for the Humanities (NEH) y la Biblioteca del Congreso (LC). El NDNP es un esfuerzo a largo plazo para ofrecer acceso permanente a la colección digital nacional de información bibliográfica de la prensa y periódicos históricos seleccionados, digitalizados por los beneficiarios del NEH en todos los estados y territorios estadounidenses. El programa le proporciona a la Biblioteca del Congreso un terreno de prueba para el desarrollo de programas de digitalización distribuidos a gran escala y para la predicción de necesidades a largo plazo para el manejo y la preservación de los recursos digitales. El desarrollo se centra en crear sustitutos de páginas de periódicos digitalizadas a través de un esfuerzo distribuido, realizar la ingestión de los objetos digitales resultantes a un sistema, brindar acceso a los datos de manera amigable al usuario, al mismo tiempo que implementa un sistema capaz de sostener el contenido para su uso futuro.

Los objetivos fundamentales del programa son, a largo plazo, proveer acceso mejorado a periódicos seleccionados mediante la creación y agregación de millones de páginas digitalizadas de comunidades geográficamente diversas, mientras se replantea el propósito de los datos bibliográficos y de existencias de más de 140.000 títulos estadounidenses en un sistema de libre acceso y búsqueda. En virtud de que la colección nacional de prensa de los Estados Unidos está dispersa entre cientos de bibliotecas en todo el país, se adoptó un modelo de selección descentralizada y conversión digital con agregación de datos suministrado por la Biblioteca del Congreso para acceso y preservación.

Un acuerdo entre el NEH y la LC establece claramente las responsabilidades de las dos agencias en el desarrollo del programa nacional global. Mientras que el NEH administra y financia competencias de premios anuales entre instituciones a escala estadal para seleccionar y convertir periódicos históricos a forma digital, la LC se centra en las especificaciones técnicas del programa, el manejo de datos y ofrecer el contenido al público. Las instituciones de nivel estadal, conocidas en el programa como "beneficiarias", son responsables de la selección de los periódicos publicados en su estado, de acuerdo con los lineamientos del programa y su conversión a la forma digital válida para su agregación central en la LC.

El desarrollo de las nuevas tecnologías de digitalización, el reconocimiento de texto y los motores de búsqueda habilitan al NDNP para brindar un acceso mejorado y el descubrimiento de este material, así como para alcanzar el liderazgo necesario para establecer las mejores prácticas y normas de digitalización y estructura para la prensa histórica destinada a convertirse en un recurso electrónico sostenible.

El entorno de gestión de datos del NDNP se basa en los requisitos que respaldan los cuatro flujos de trabajo principales según están identificados en el Modelo de Referencia de los Sistemas Abiertos de Información de Archivos (OAIS): ingestión, archivo, diseminación y gestión de preservación. Desde el inicio, la Biblioteca del Congreso reconoció el alcance del programa previsto – millones de páginas de periódicos producidas por muchas organizaciones diferentes durante varias décadas (equivalentes, por lo menos, a cientos de terabites). El compromiso entre las agencias que reciben financiamiento público para manejar el acceso a este contenido requirió hacer énfasis en la creación de activos digitales de acuerdo con las normas emergentes y las mejores prácticas uniformes.

Los datos bien formados que operen dentro una infraestructura técnica robusta serían el mejor enfoque para asegurar una gestión con efectividad de costos del contenido a largo plazo.

Estas decisiones técnicas también estuvieron determinadas por las realidades de la estructura del programa completo:

– El contenido – versiones analógicas (microfilme, papel) de los periódicos históricos estadounidenses reside principalmente en repositorios de los estados, en lugar de una biblioteca nacional, y por lo tanto el programa requiere una producción distribuida de los activos digitales.
– El financiamiento para aplicar nuevas tecnologías para mejorar el acceso a este material es finito, por lo tanto,
  • el contenido incluido en el programa será selectivo, en lugar del corpus completo disponible;
  • los requerimientos técnicos para los materiales convertidos deben abarcar la potencial reutilización y reprocesamiento a largo plazo (escanear una vez, usar muchas veces);
  • El programa debe ser un modelo para otros esfuerzos que puedan interoperar en el futuro – compartir las mejores prácticas, especificaciones de conversión y estandarización del acceso básico a los periódicos históricos.

– Demostración del buen uso de los fondos públicos para proveer acceso abierto y permanente;
– Con la expectativa de cambios, se debe evitar cerrar opciones, mediante el desarrollo de un entorno de sistema abierto, expandible y modular.

A partir de su experiencia con la digitalización a gran escala de materiales históricos, la LC desarrolló las especificaciones técnicas para el contenido del NDNP con base en las mejores prácticas existentes.

La interfaz del usuario con los datos del NDNP, conocida como *Chronicling America: Historic American Newspapers* (http://chroniclingamerica.loc.gov), es de libre acceso para el público y está disponible en el sitio web de la Biblioteca del Congreso. El sitio actualmente incluye un aproximado de 5 millones de páginas de periódicos de más de 700 títulos, publicados entre 1836 y 1922 de 25 estados y el Distrito de Columbia, con 3 estados adicionales pendientes por incluir.

Un componente importante en el desempeño del papel de la Biblioteca del Congreso en este programa es el desarrollo de un entorno de sistema que asegure que los activos digitales adquiridos de distintas fuentes durante un período prolongado de tiempo sean sostenibles. El entorno debe garantizar que cuando las personas, los procesos y las tecnologías cambien, se pueda acceder al activo digital (de manera transparente y automática, si es posible) para su uso y, potencialmente, su transformación. La arquitectura adecuada de los repositorios es un componente esencial para determinar si un entorno de preservación digital es exitoso.

Conjuntamente, la LC, el NEH y los beneficiarios participantes han desarrollado una fuerte comunidad de práctica alrededor de la digitalización de los periódicos históricos que:

- permite un cuerpo uniforme de contenido creado por productores múltiples,
- satisface las necesidades básicas de acceso de los usuarios más allá de las versiones analógicas,
- tiene un alto grado de sostenibilidad y
- mejora el acceso a los periódicos estadounidenses en general.

Saving Our Past into the Future: the Preservation and Digitisation of Old Newspapers at Shanghai Library

by Chen Xuyan, Senior Coordinator, International Relations, Shanghai Library, China

The Old Newspaper Collection of Shanghai Library

An Introduction

Being witness to the history, the old newspapers are regarded, especially for decades or hundreds of years after their publication, as key resources and irreplaceable documents which provide vivid accounts of the places, events and people. A wide variety of information has been published on newspapers, which was never intended by their producers to be a permanent means of storing information for researches of all kinds in the future.

Shanghai Library, the second largest library in China next to the National Library in Beijing, boasts one of the finest old newspaper collections in the country. Some of the holdings are quite unique and even exclusive, including, among others, the *North China Daily News*, the earliest English language newspaper appeared in Shanghai; *The Shanghai Xinbao*, the first Chinese language newspaper under the foreign ownership in the city; and *Shen Bao* (transliterated as *Shun Pao*), the most influential Chinese language newspaper in the modern and contemporary history of the Chinese journalism.

People of all walks of life have been turning a lot to these old newspapers for information, and even for inspiration. In the mid 1990s the screenwriters of a TV opera crew spent months at Shanghai Library exploring the archives of *Shen Bao*. The newspaper involved itself during the late Qing Dynasty in the coverage of a celebrated wrongful conviction case of "Yang Naiwu and Xiao Bai Cai", and successfully helped mobilise the elite opinion and public sentiment in favor of the defendant Yang Naiwu, who was believed innocent and finally exonerated as a result of the intervention from the Imperial throne. The TV series produced by the same name as the case became a much watched and talked about one, and is still frequently quoted by the authorities as need for vigilance against corruption and miscarriage of justice.

Today Shanghai Library is the one and only library in the world that keeps the paper's all original issues from April 30, 1872 until May 27, 1949.

Libraries Specialised in Newspaper Collection in Shanghai – The Early History

The first library in Shanghai was built in 1847 as a repository of the French Jesuit mission, known as the Library of the Catholic Church of Xujiahui or Bibliotheca Zi-Ka-Wei[1]. In the next 100 years its collection grew to over 100,000 titles in 200,000 volumes – 80,000 volumes in different European languages and 120,000 in Chinese, many are pre-1800 rare editions, together with the extensive holdings of local gazetteers and modern newspapers published in China, including a complete collection of *Shen Bao*'s issues. In November 1955 it was taken over by the Shanghai Municipal People's Government, and reopened in January 1957 as a branch of the Shanghai Library.

1. The Bibliotheca Zi-Ka-Wei.

2. The reading room of the Bibliotheca Zi-Ka-Wei.

The Jiazishe Library was established in 1924 with a collection emphasis on the current newspapers and back issues of major old papers. From 1929 it began to edit the content indexes of daily papers and periodicals. The Library was renamed as Hung-Ying Library after receiving the sponsorship from Mr. Ye Hung-ying, a capable local businessman, in June of 1933. By 1946, it had built up a collection of 136 daily papers, 150,453 volumes of books and journals, 181,366 newspaper clippings, as well as some 300,000 newspaper-and-journal index cards. In January 1955, it merged with the fledging News Library to become a new Shanghai Newspaper Library.

1. Zi-Ka-Wei is the Chinese Wu or Shanghai Dialect pronunciation for Xujiahui.

As of the end of 1954, the Shanghai Newspaper Library had an inventory of 438,125 books, 336,274 volumes of magazines, 1,151 bound journals and 7,742 bound newspaper files. In 1953 the Library issued the first publication of *The National Index to Major Chinese Periodicals*, which was renamed *The National Index to Major Chinese Newspapers & Periodicals* in 1956, as the earliest searching tool to retrieve newspaper titles and articles.

Then finally in October 1958, the Shanghai Newspaper Library, together with the Shanghai Municipal Library of Science and Technology and the Shanghai Municipal Library of Historical Documents, were merged with the Shanghai Municipal Library, founded on July 22, 1952 under the name of "Shanghai Library", in succession to its acquisition of the Bibliotheca Zi-Ka-Wei a year earlier.

Featured Holdings

According to the *Catalogue of the Chinese Newspaper Collection of* Shanghai Library published in 1982, the Library is home to altogether 3,543 titles of Chinese language newspapers published daily, weekly, or at other regular intervals between 1862 and 1949, spanning from the late Qing Dynasty to the Republic of China Period.

The most valuable newspaper collection is no doubt the complete runs of *Shen Bao*, romanized as *Shun Pao* and known in English as Shanghai News. It was founded in April 1872 by a British businessman Ernest Major, yet the emphasis was laid on news and issues of interest to Chinese. It was in publication for 78 years in 25,600 issues until May 1949, not only the longest, but also the most influential in all extant or discontinued Chinese newspapers. It made its reputation by reporting almost all the major events during this period of history and voicing its distinctive opinions through editorials.

Another rare early newspaper was *Shanghai Xinbao*, founded in November 1861 as a weekly by the American Methodist missionary Young John Allen. It is the first Chinese language newspaper published in Shanghai. A distinctive feature was its comprehensive coverage on the *Taiping Heavenly Kingdom*[2], which was later compiled and published by Shanghai Library in 1964. The paper lost its business competition with *Shen Bao* and then stopped publication on December 31, 1872. This collection is the most complete and the best-preserved that cannot be found elsewhere in China.

Other important and hard-to-acquire Chinese newspapers held by Shanghai Library include the *Hui Pao* (1874, in 66 issues), *Yi Pao* (1875, in 218 issues) and *Xin Pao* (1876, in 116 issues). A second catalogue, produced by the branch of Bibliotheca Zi-Ka-Wei, records a total of 92 titles of different foreign language old newspapers in more than 1,000 volumes that were published from 1851 to 1949. Of the papers, 66 titles were published in European languages, including English, German, French, Spanish, Portuguese, Italian and Russian; the other 26 titles are in Japanese.

Of all the early foreign newspapers, there is *North China Herald*, which is known as the first newspaper in Shanghai. It was first published on August 3, 1850 as an English language weekly and the gazette of the British Supreme Court for China and Japan and the British Consulate. A daily edition commenced on June 1, 1864 as the *North China Daily News*. It was so influential – in fact the most in Shanghai and even throughout China, nicknamed "the Chinese version of *The Times*" – that everything from the public notices of the British Consulate and the then Shanghai Municipal Council to the loss statement by a small time merchant, wouldn't come into effect until it was published on the paper. The *North-China Herald* and the daily

2. The *Taiping Heavenly Kingdom* is a rebellious state between 1851 and 1864 in South China.

3. (Left) The August 15, 1945, *Shen Bao* headline with the surrender edict by then Japanese Emperor Showa; (right) A *Shen Bao* supplement on July 12, 1938.

edition suspended publication after December 8, 1941 during the Pacific war. Publication of the *Herald* was never resumed. On March 31, 1951, the *North China Daily News* suspended its 101-year-long history of publication.
Other survived rare newspapers in various western languages include *China Press, Shanghai Times, Shanghai Evening Post and Mercury, Shanghai Jewish Chronicle, Ostasiatische Lloyd, Journal de Shanghai, Echo de Chine*, to name a few.

Preservation of Old Newspapers

Storage Conditions

The Chinese old newspapers are now stored in the closed stacks on the eighth floor of the main building of Shanghai Library, which was put into use in 1995; and the foreign language old papers are housed on the fourth floor of the Bibliotheca Zi-Ka-Wei. Both buildings are quite new in terms of the duration in service and are in good form – the Bibliotheca Zi-Ka-Wei, though founded in the mid 19th century, was just overhauled in 2002-2003 to be renovated into a more comfortable and safer place for both readers and the collections, while maintaining its original Vatican architectural style.
The Shanghai Library was designed and constructed in full conformity to the National Standard of Library Building JGJ38-99. Temperature of the special collection storages, where the old newspapers are housed, are kept at 24°C to 25°C in summer and 12°C to 15°C in winter, while the relative humidity constantly at about 55%, within 10% of daily change, through separate air-conditioning systems. When power is out, lime or charcoal is used to absorb moisture. There are also air purification devices installed in the microfilm collection stack to prevent the entry of hazardous gas and dust.

4. Air-conditioning system to keep the constant temperature and humidity.

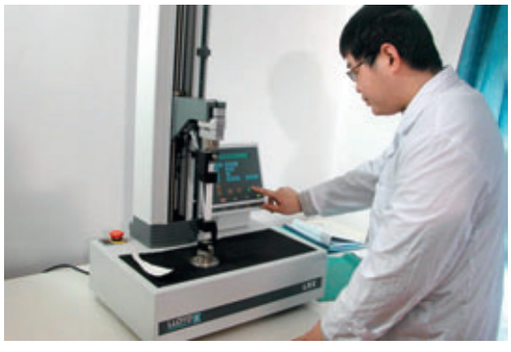
5. Air purification facility installed in the special collection stacks.

The original old newspapers are stored unfolded on stainless steel closed shelves. Temperature and moisture recorders are installed. The foreign language old newspaper however, are held in a semi-open storage space of the Bibliotheca Zi-Ka-Wei, so the stack is installed with shutters and a ultraviolet filtered, energy-saving lighting system to minimise the sunlight and UV exposure that may cause damage to the newspapers.

Conservation Treatments

The Department of Collection Preservation and Conservation is a laboratory established by Shanghai Library to experiment the practical techniques and solutions for preservation of library collection, especially the old and special materials. Thanks to the financial support given by the city government, which earmarks an annual fund of 1 million Yuan to rescue the ancient and rare literatures, the Department managed to equip the laboratory with refined facilities and professional conservators. A number of research and pilot projects have been conducted over the recent years, making achievements in for instance repairing and fixing the thread-bound ancient books, preventing books and journals from mould, and applying preservative processing to the document repairing pulp.
The laboratory has worked out a practice of dust removal and pest elimination. A self-made pyrethrins agent is applied as pesticide to the special collection stacks twice a year, generally when spring turns into summer, and when autumn is replaced by winter, during which the circulation has never been interrupted. Chemical papers containing the same ingredients of the pesticide are inserted between pages of books and newspapers to protect them from the invasion and damage of silverfish. The recent application of an ultra high-frequency pest killing devices, designed to phase out the traditional chemical fumigation which is deemed as a threat to the environment and people's health, has turned out to be equally effective.

6. Staff spraying the home-made pesticide in the ancient document stacks.



7. Staff working on the Lloyd LRX to test the torque force of the paper.

The laboratory also carries out a thorough examination every three months to evaluate the state-of-being of the library holdings, gathering data in terms of the type, quantity, grade, location, storage space, raw materials and the collection degradation to detect problems and therefore to make priorities for remedies.

In 2010 the laboratory added some new facilities, including paper fiber tester, paper tensile force tester and acidification detector, to measure the properties of the paper collections.

Access Policies

Many of the old newspapers held by Shanghai Library are now more than 70 years old, which are becoming increasingly fragile as result of natural aging or heavy use. Thus restrictions have to be imposed on the access to the original copies.

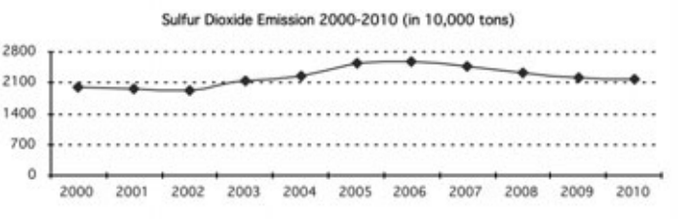
Amounting to 80% of the Chinese old newspapers have already been microfilmed in the 1980s. Now the 6,000 plus reels of 35mm microfilms are browsed primarily on the microfilm readers, and there is no access allowed to the original copies. Access to the remaining newspaper titles without being microfilmed can be offered only in the present of a "letter of introduction" which is issued by the employer or other authorities of the user.

Reading service to the original issues of the western language old newspapers in the Bibliotheca Zi-Ka-wei is also restricted to a substantial extent. No permission is granted to access the only extant original copies and the duplicates that are severely outworn or fragile. Xeroxing and scanning, as believed to cause direct damage to the papers, are prohibited. Photocopies of the originals, though very rarely accessible in the market, have all been purchased – such as that of the complete back issues of the 1832 founded *Chinese Repository* – to meet the growing reading demand.

Challenges

There are a number of factors attributed to the deterioration of the old collections including the newspaper, but the most fatal one is the acidification of paper. The early newspapers published from the late 19th century to the mid of 20th century in Shanghai and the rest of China, were basically printed on newsprints produced by cheap and high-yielding papermaking techniques introduced from Europe, whose outputs were in quite low pH values.

Unlike the Xuanzhi, or Xuan Paper made with traditional Chinese papermaking raw materials and techniques, which is believed able to sustain for as long as 1,000 years with great resistance to crease, corrosion, light and mold, the paper produced at the turn of the 20th century in China has a much shorter life for no more than 100 years. A latest research conducted by the National Library of China, *The Survey and Analysis on Acidification of the Paper Collections and Preservation Situation*, has found a sharp pH decrease with paper produced from the late Qing to the early Minguo (Republic of China, 1911-1949) period, when the majority of old newspaper holdings of libraries in China were published. The pH values of the paper-based literatures of the 1920-1930s were far below 4.0, and didn't rise to above 6.0 until the 1960s when technical progress were gradually made for both paper manufacturing and collection preservation.



8. The SO_2 emission in China (2000-2010).

A major extrinsic factor that speeds the paper degradation is the growing air pollution, a by-product of the industrialisation over the recent decades in China. The paper, even made in neutral or alkaline, is subject to absorbing the environmental pollutants, especially the acidic sulfur dioxide, which finally forms sulfuric acid that directly acidifies paper, and turns paper yellowish and brittle. In 2006 China's record high sulfur dioxide pollution was the worst in the world.

Saving the old collections is no easy task. The no.1 issue is the library funds. The Preservation and Conservation Laboratory is well funded and equipped, yet not sufficient enough. According to a specialised de-acidification solution supplier, the de-acidification and stabilising process for per 120 tons of paper collection costs approximately USD 4 million and 48 hours, plus a 200 sq meter of space to load the equipments and another 100 sq meter to rearrange the materials. For public libraries around the world that are running on shoestring budgets today, with no exception in China, this is simply not affordable.

Protection and Rescue Strategies

Protection of the old newspapers, along with other treasured holdings, has always been a priority of Shanghai Library's collection policies. The necessity of preservation has already been envisioned in the early 1950s when the first efforts were made simultaneously.

Providing Searching Tools

Searching articles by title, author, date, or keywords by a comprehensive newspaper index proves effective in minimising aimless paging through the original newspaper copies, which leads to evitable but serious damages, while facilitating in locating to the desired contents.

Publication of the monthly *National Index to Major Chinese Newspapers* commenced in March 1955. The first issue collected 5,500 entries from 73 newspaper and periodical titles. The *Index* was discontinued from October 1966 for 7 years, and resumed publication in October 1973 under the name of the *National Index to Chinese Newspapers & Periodicals*. In 1980 the *Index* was divided into the "Natural Science" and the "Social Science" volumes. And from 1995 onwards, it has been issued in both print and the CD-ROM versions. Since 2004 the *Index* is also available as online databases.

Today the *Index* has developed into a comprehensive database in print, disc and online editions, encompassing some 15,000 Chinese language newspapers and periodicals published in China (including Hong Kong, Macao and Taiwan), with entries accumulated to more than 30 million in total, and the earliest dates back to 1833.

Preservation Microfilming

Preservation microfilming is an established strategy for producing copies of paper collections. A petty attempt had been made much earlier at Shanghai Library – the Shanghai Newspaper Library began the experiment of replicating the historical materials in microform by photography in 1956. A small team was organised in 1958 at Shanghai Library, after the incorporation of the Newspaper Library, to provide document copying and enlarging for reading purpose. Copy by photography and Xerox was practiced on some back issues of *Shanghai Xinbao* over the 1960s.

The extensive microfilming was initiated in April 1983, when 11 major Chinese libraries were designated as the first libraries to carry out microfilming on paper collections at a workshop convened by the Division of Library of the Chinese Ministry of Culture. In January 1985, the China National Microfilming Center for Library Resources was established in the National Library of China to serve as headquarter of the nationwide microfilming projects. Microfilming of the modern newspapers and journals was assigned to Shanghai Library, which was conducted at the Bibliotheca Zi-Ka-Wei where most old newspapers were housed.

At the moment, many of the newspapers published in modern times had become badly outworn after heavy use. What's more, some collections were often incomplete in back issue collections due to their irregular and prolonged publication spans. Missing of certain issues or pages was detected to a different extent with many paper titles. So the first step was to regain the completeness of the newspaper collections.

Working groups were sent to 18 libraries and newspaper depository units throughout the country in acquisition of the lost original issues. Substantial progress was made: altogether 691,786 pages in 18,412 issues from 275 newspaper titles, amounting to nearly one third of the total estimated microfilming workload, were acquired to make up for the lost originals.

The preservation microfilming was carried out in strict accordance with the national standards stipulated by the National Standardisation of Microphotography Technology Committee. Fragile newspapers were repaired and stabilised before being filmed. Testing was compulsory for each filming to ensure the quality of the master negative. As of the end of 1989, a total of 264 million pages of 500 plus Chinese language newspaper titles published before 1949, involving almost all the Library's Chinese old newspaper holdings, have been reformatted to more than 6,000 reels of 35 mm microfilms.

From the 1990s onwards, microfilming at Shanghai Library has mainly been oriented to the "Minguo Periodicals" (periodicals published in 1911-1949), and is no longer applied to reformatting the newspapers. The replication service based on the existing printing negatives however, is available on demand.

The service positives of the microfilmed newspapers can now be viewed on the microfilm readers, requiring a visit to Shanghai Library. Printing from the microfilms is offered on a nominal service fee.

Access in Digital Form

Shanghai library has purchased commercial e-newspaper databases involving altogether 3,442 digitized newspapers (1,896 Chinese newspaper and 1,546 in foreign languages, including many of the titles that are held by the Library). Less than half of these e-papers (1,438 titles) are offered in text edition only (without images and advertisements); all the rest are in full edition in PDF or html format.

Permission-based access to the e-newspaper databases is granted via the "e-card-through Services", a 24*7 online remote service platform (http://eservice.digilib.sh.cn:8080/arrayLogin/index.jsp) customised by Shanghai Library for its registered users. Visitors can also view some 1,300 titles of digitised newspaper via LAN on the computers installed in the Library.

E-newspapers are also acquired at no charge through donation. The offices of the *Xinmin Evening, Wenhui Daily* and *People's Daily*, three most circulated current newspapers in China, have bestowed the complete runs in CD-ROMs from the early 1930s till 2005 at the latest. A web page (http://newspaper.digilib.sh.cn/website/index.asp) titled *Shanghai Library e-Newspaper Guide* was also created to incorporate links to 498 online newspaper web sites so that access to certain newspaper collections is further diverted. The page is regularly updated.

Shanghai Library's Newspaper Digitisation Project

Degradation of the original old paper-based collection is actually irreversible. The microfilms, on which conservators had previously laid much hope for safe storage for an expected long period, are also found susceptible to light, pollution or other environmental impacts. Despite the cost concern and a number of technical considerations such as the size and brittleness of the newspaper pages, digitisation is, though far from being widely accepted at the time being, an effective preservation technique to protect the decaying library materials.

Digital Scanning

The first digitisation projects were conducted in 2002 to rescue some selected ancient rare books and Shanghai Library's featured special holding, the genealogies. Digitisation production lines using the most sophisticated equipments and techniques, combining with a series of GMPs were built up over the following years. The digitisation center of Shanghai Library has been scanning up to 5 million digital images every year from the original newspaper pages, periodicals and manuscripts.

What are being digitised currently at Shanghai Library are the *North China Daily News*, the *North China Herald*, and a collection of selected Chinese newspapers published in 1950-1960, which were not microfilmed in the 1980s when they were still young, but are now badly damaged largely due to the poor quality and acidic newsprints. The scanning center uses hardware such as Zeutschel A0 and OS 10000 A1 scanners and the Minolta Ps 7000 scanners to capture images of newspaper pages above A1 in size, in contact or non-contact manner; and applies Photoshop software and SilverFast plug-in for scanning and image processing.

According to the counting at the end of 2007, there are altogether 600,000 pages of the *North China Daily News* and the *Herald* weekly to be scanned. As of the end of 2011, digitisation of more than half of them has been done, with 60,000 pages digitised in 2011 alone. It was anticipated to take another two years to complete the digital scanning as well as creation of all metadata and development of an automatic indexing system.

The digitising workload of the 1950-1960 newspapers is difficult to be estimated, since the Library is yet to have an accurate statistics on the papers whose properties are unsustainable. The yearly output over the past three years averages 100,000 pages. The general catalogue of the digitised 1950-1960 newspaper won't be available until the scanning work is finalised.

Newspaper scanning is conducted in compliance with the standards and guidelines stipulated by the National Digital Library Project, as well as Shanghai Library's own "good manufacturing practices". Instead of going through an Optical Character Recognition (OCR) process that converts images into texts, it yields high fidelity images in 8-bit grayscale at 300dpi. Scanning is operated on duplicates, rather than on the original copies for protection purpose. When the only copy to be scanned is broken or wrecked, a white paper should be placed underneath and, when there are wrinkles and folds, ironing beforehand is strictly asked. Joints between every two pages are required to be fully included in the scanning area. Like the "testing" step in microfilming, a "preview" procedure is mandatory before each operation to optimise the brightness, contrast and alignment to fit the various conditions, such as the over-or-under inked texts.

Images created from the scanning process are stored at the end of each day as tagged image file format (TIFF) in the movable disks as master for long-term archive, and then returned to the DAT tapes. A three-level file name is simultaneously given to each image, with call number and newspaper title at the first level, the four-digit publication year and two-digit issuing month at the second and the number of page at the third, e.g. "G10.Workers Daily\194908\01-01.tif".

The file name list is then loaded onto Shanghai Library's library management system to generate the metadata of title, date, number of pages, and other elements. The metadata encoding system of Shanghai Library went through a transformation from MARC to the DC (Dublin Core) schema in 2004, with the first success of applying DC metadata elements to the description and index system for Shanghai Library's manuscript collections in 2005. Then the DC schema was gradually adopted for description of resources of all other kinds.

Digital Conversion from Microfilms

The first digital conversion work in China was done by the National Library of China in 1998, primarily targeted to the books and journals published in the Republic of China period, at an annual output of 1.6 million pages. No newspaper titles were considered in this project. Elsewhere in China, some big public libraries in Guangdong, Shanxi, Zhejiang and Shandong province have purchased digital conversion facilities to carry out pilot projects, yet far from being capable of manufacturing digitised surrogates massively.

Scanning newspapers from the existing 35mm microfilms has not yet started at Shanghai Library, but the equipments and a five-year master plan has already been put into place. To be converted will be newspapers and periodicals that were microfilmed in the 1980s. The work will be kicked off in 2013, with estimated 1.5 million pages to be digitised every year.

The Web-based Access to the Digitised Collections

Generation of the metadata of the scanned 1950-1960 Chinese newspapers and the *North China Daily News* and *North China Herald* is not completed yet. So Shanghai Library developed a web-based interim system, on which the digitised papers can be viewed, printed and searched by date (see figures below), by means of extracting and copying the information of newspaper title and date of issue from the file directories and file names into the system registry table.

9. The landing page of the web-based access to the digitised *North China Daily News*.

10. The viewing interface of the digitised *North China Daily News*, whose pages can be flipped over and printed out.

According to the technical roadmap, the metadata of the digitised newspapers, as soon as they are created, will be loaded to Shanghai Library's LIS system. The TIFF files will be converted into PDF format for full text browsing and retrieval functions. A connection will then be set up to match each PDF file with the corresponding metadata to ensure the appropriate access. All these will then be loaded into a more sophisticated unified platform which provides access to ideally all the historical collections held and digitised by Shanghai Library.

The Unified Platform for Historical Resources

The Unified Platform for Historical Resources went live in September 2010 after three years development to provide one-stop retrieval, reading and printing service. Access is now given to 7 digitised historical resource databases including the an-

cient rare books, genealogy ar chives, SHENG Xuanhuai [3] Archives, the Minguo (Republic of China period) books, the Minguo periodicals and dissertations, the old Japanese books, the old western language books (and in the near futur e, the old newspapers). The platform is now accessible on mor e than 30 computers in 3 ancient and modern document reading rooms. The platform features smart retrieval in single, multiple, or overall databases by title, author, keyword, call number or subjects. Search can be done in a pr ecise or fuzzy manner, and the precise retrieval can be fulfi lled by up to 5 keywor ds. Inputs of traditional and simplifi ed Chinese characters ar e both r ecognizable. The *Chinese Library Classification* and *Chinese Ancient Document Classification* are available r espectively for the ancient book database and the Minguo book database. Resources in full text can be viewed and printed in PDF with contr olled permission. Users' data are logged in full details.

The platform is loaded with 2.79 million metadata, with r elation to 39.098 million pages of 2.371 million articles. Statistics in September 2011 showed that since its inception a year before, the Platform had pr ovided 106,927 retrievals, 2,265,443 pages of full text view and 4,490 print-out services.

Conclusion

Shanghai Library has been engaging itself in providing different users with wider access to its featured collections, by exploring new and sustainable ways of pr eservation and conservation. Progresses have been made, yet as time goes by, new problems are emerging. The micr ofilms for instance, once believed an endurable media, have pr oven to be sensitive and vulnerable. At the time being, digitisation is no doubt a r eliable option to protect the library collections, yet we still need to understand a variety of questions fr om the economic, technological, legal and organizational perspectives – this may include the appr opriate storage and maintenance of the digital data, the costs for hardware and software, the IPR concerns if a consortium is set up to share the outcomes. The answers, which require theoretical and practical verifi cation, will help us to guarantee that library digitisation projects will be carried out tomorr ow more effectively and less expensively, to hopefully save our treasured heritage well into the future.

Acknowledgement

I would like to extend my heartfelt thanks to Mr . WANG Renfang, Mr. FAN Zhaoming, Mr. LIN He, Ms. XIA Cuijuan, Mr. XIA Hai, who have given me much resource, guidance and inspiration. My special gratitude also goes to Dr. WU Jianzhong, the Director of Shanghai Library, who has of fered a lot of convenience and encouragement all the way.

11. The landing page of the Unified Platform for Historical Resources.

12. Viewing page of PDF-based full page on the Unified Platform.

3. SHENG Xuanhuai (1844-1916) was a celebrated entrepreneur of the Qing Dynasty. He was also the Minister of T ransportation, and founder of some early higher education institutions, banks and the Red Cross Society of China. His r esidence in Shanghai adjoins the curr ent premises of Shanghai Library.

References

1. Wu, Min. 2004. "Newspaper in Shanghai Library". Paper Presented at 20th Business meeting of the IFLA Newspapers Section at Shanghai Library, Shanghai, China, March 29, 2004.

2. Feng, Jieyin, Shi, Zhonghua and W u, Zhongxia. 2005. "Preserving our collection – the new building of the Shanghai Library". World Library and Information Congress: 71th IFLA General Conference and Council, "Libraries – a Voyage of Discovery", August 14-18, 2005, Oslo, Norway.

3. Li, Chunming, and Zhang, W ei. 2006. "Microfilming and Digitization of Newspapers in China". Pr e-conference of WLIC 2006 Preservation and Conservation in Asia, National Diet Library, Tokyo, Japan, August 16-17, 2006.

4. Zhang, Wei. 2007. "Preservation and Protection of Newspapers on Research Microfilming & Digitization". In *Library Development*, Vol. 79, No. 2, 37-40.

5. Sangeeta, Keisham. 2006. "Digitisation of Newspaper: An Easy Access to Information". 4 th Convention PLANNER -2006, Mizoram University, Aizawl, November 9-10, 2006.

6. Mieczkowska, Suzanne. 2002. "Digitised Newspapers at Norfolk and Norwich Millennium Library". In *Collection Building*, Vol. 21, Iss: 4, pp. 155-160.

7. Office of *Shanghai Chronicles*. Online version, Shanghai Chronicles: *Libraries in Shanghai*. http://www.shtong.gov.cn/node2/node2245/node4457/node55856/index.html (Accessed March 1, 2012)

8. The British Library. *About the British Newspaper Archive*. http://www.britishnewspaperarchive.co.uk/help/about (Accessed March 3, 2012)

9. The Library of Congress. *Collections Care: Preservation Guidelines for Digitizing Library Materials.* http://www.loc.gov/preservation/care/scan.html (Accessed March 5, 2012)

10. The Library of Congr ess. *Collections Care: Preserving Newspapers.* http://www.loc.gov/preservation/care/newspap.html (Accessed March 5, 2012)

11. Shanghai Library. *The Ancient Document Restoration Service*. http://www.library.sh.cn/fwzn/wxxf/index.htm (Accessed March 5, 2012)

The National Library of South Africa and the Digitization of the Early Years of the Black Press

by Douwe Drijfhout, Executive Head, Preservation Services, National Library of South Africa

So much of South Africa's troubled and complex past is documented, stored and bound on newspaper pages. Newspapers are important sources of social, political and literary information. They contain a reflection of opinion and life of a given time in history that is not recorded anywhere else. A true reflection of the zeitgeist as it was experienced and seen through the eyes of contemporaries.

The National Library of South Africa originated from libraries that existed since the early and mid nineteenth century. The newspaper collections date back to the earliest newspapers published in the country. Most of them on microfilm of which many were filmed by the library itself. Both campuses in Pretoria (City of Tshwane) and Cape Town keep large collections of original print copies of which some date back to the earliest publications.

In their unique contribution towards understanding the origins of the South African Press, Switzer & Switzer (1979: vii) describe it as sectional throughout its history. Race - not language, religion or culture – proved to be the dominant characteristic of its segmentation. Thus giving the press in South Africa a unique status among the world's mass media. But also the oldest, most extensive and varied collection of indigenous serial publications of this kind in sub-Saharan Africa.

The Switzers (1979: 1-22, 1997: 3) distinguished mainly three stages in the development of the Black Press, namely:

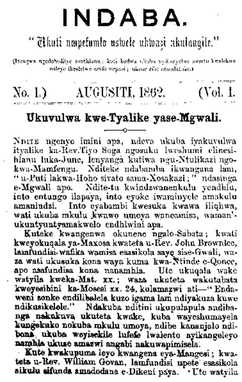
- 1830's-1880's – Christian Missionaries
- 1880's-1930's – An Independent Protest Press (the golden age)
- 1940's-1960's – From Protest to Resistance (African nationalism)

The Black Press is defined according to its readership: thus newspapers that were directed at an African, Indian, and Coloured audience. Not necessarily owned or edited by Africans. Newspapers that were often bi-lingual in English with contributions in Zulu, Xhosa, Tswana and Sotho.

The history of the press in South Africa dates back to 1800 when the Governor of the Cape initiated the publishing of the *Cape Town Gazette* and *African Advertiser*. The first private newspaper *The SA Commercial Advertiser* was published in 1824.

1830's-1880's – Christian Missionaries

The first serial publications in the vernacular languages of South Africa were printed on presses operated by Christian missionary stations around the country. These were religious tracts focussing on devotional and evangelical matters. Controlled by the missionaries but written and later edited largely by Africans. The first serial was published in Tswana in Kuruman in 1836: *Morisa oa Molemo* (Shepherd the Good). The earliest known African newspaper in South Africa *Umshumayeli Wendaba* (Publisher of the News) was published in Grahamstown in 1837. *Indaba* (1862) was the first bi-lingual newspaper in both English and Xhosa. Published by African teachers and students at Lovedale.


INDABA.

"Ukuti umpefumlo uswele ulwazi akulungile."

([illegible])

No. 1.) AUGUSITI, 1862. (Vol. 1.

Ukuvulwa kwe-Tyalike yase-Mgwali.

NDITE ngenye imini apa, ndeva ukuba iyakuvulwa ityalike ka-Rev. Tiyo Soga ngosuku lweshumi elinesihlanu luka-June, lenyanga kutiwa ngu-Ntulikazi ngokwa-Mamfengu. Nditeke ndabamba ikwangana lam, "u-Puti lakwa-Hoho sivato sama-Xosakazi;" ndasinga e-Mgwali apo. Ndite-tu kwindzwanenkulu yendlu, into entungo ilapaya, into eyoke iwaminyele amakulu amanindzi. Into eyabambi kwesuka kwawa iliqhwa, wati ukuba mkulu kwawo umoya wancamisa, waman'-ukuntyuntyamakwelo endhlwini apa.

Kuteke kwangenwa okunene ngalo-Sabata; kwati kweyokuqala ya-Maxosa kwateta u-Rev. John Brownlee, lamfundisi wafika wamisa esasikola saye sise-Gwali, waza wati ukusuka kona waya kuma kwa-Ntinde e-Qonce, apo asafundisa kona nanamhla. Ute ukuqala wake watyila kweka-Mat. xx.; waza ukuteta wakutabata kweyesibini ka-Mosesi xx. 24, kulamazwi ati—"Endaweni zonke endilibalela kuzo igama lam ndiyakuza kuwe ndikusikelele." Ndakuba nditini ukupulapula andibanga nakukuva ukuteta kwake, kuba wayeshumayela kungekuko nokuba mkulu umoya, ndibe kananjalo ndibona ukuba weyisekile lufefe lwalento ayikangeleyo namhla ukuse amazwi angabi nakuwapimisela.

Kute kwakupuma leyo kwangena eya-Mangesi; kwateta u-Rev. William Govan, lamfundisi upete esasikola sikulu sifunda amadodana e-Dikeni paya. Ute watyila

1. *Indaba.*

Two significant mission publishing centres emerged.

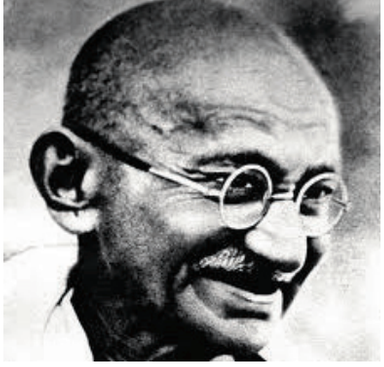
The Presbyterians, at Lovedale in the Eastern Cape and the Paris Evangelical Missionary Society at Morija in Basutuland (Lesotho). The Lovedale Mission Press printed *Isigidimi Sama Xosa* in 1873: the first African newspaper to be edited by Africans. The most important periodical produced in the first 60 years of the mission press in South Africa. Editor Elijah Makiwane was succeeded by John Tengo Jabavu in 1881. Jabavu emerged as the most prominent African political figure in the Cape colony. Some of the earliest protest poetry can be found in Isigidimi.

The Paris Evangelical Missionary Society established a press at Morija in the present day Lesotho. The Morija Printing Works eventually printed in up to 45 African languages for various countries in sub-Saharan Africa, including newspapers. Morija played a major role in the development of the Sotho language and literature.

1880's-1930's – An Independent Protest Press (The Golden Age)

From 1880 the mission presses no longer had a captive audience. A new generation of Mission educated activists emerged that marked the beginnings of an indigenous black literary tradition dominated by African nationalist newspapers. Mobilizing African opinion against segregation and discrimination. The first African political bodies were established. Literacy rates among Africans rose slowly from an estimated 6.8% of the adult population in 1911 to 9.7% in 1921. An important theme was the role of Africans in the newly established Union of South Africa (1910).



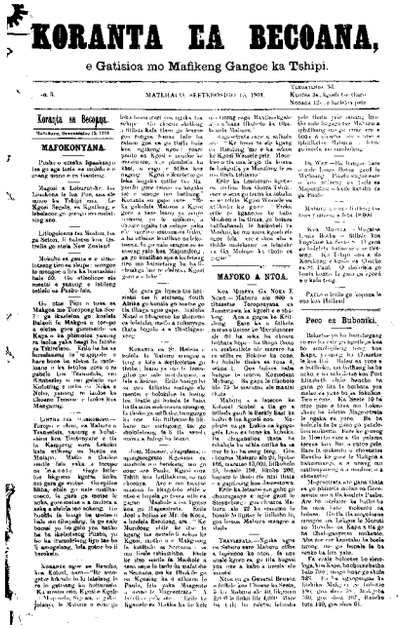




2. From left: John Tengo Jabavu and his son Davidson Don Tengo, around 1903; Walter Rubusana; Solomon Plaatje;. Mahatma Gandhi; Abdullah Abdurahman.

In the Eastern Cape John Jabavu established *Imvo Zabatsundu* (Black Opinion) in 1884. The newspaper was funded by so-called 'friends of the natives' in Cape Parliament. It was the first black owned and controlled newspaper in South Africa. Jabavu however sided with the Afrikaner Bond. He urged the British to adopt a policy of reconciliation with the Afrikaners. African opinion was divided along white party political lines and the resulting South African (Anglo-Boer) War. *Imvo* being perceived to be an organ of the Afrikaner Bond, Jabavu's leadership was questioned.

Izwi Labantu was launched in 1897 with financial backing of Cecil John Rhodes, under the leadership of Walter Rubusana, founder-member of the South African National Congress established in 1912. Other parts of the country followed with *Koranta ea Becoana* (1901) by Solomon Plaatje, *Ilanga lase Natal* (1903) by John Dube, *Indian Opinion* (1903) by Mahatma Gandhi and *APO* (African Peoples Organisation) *newspaper* (1909) by Abdullah Abdurahman.

Plaatje played a major role in African nationalist politics in the early years of the South African Native National Congress. *Koranta ea Becoana* was the first Tswana-owned newspaper in Southern Africa.
Dube was the first president-general of the South African Native National Congress (which later became the ANC). Indian Opinion played a significant role throughout the course of the passive resistance campaign. The *APO newspaper* was regarded as the authentic voice of the Coloured people. The APO voiced the sentiment that 'it is not race or colour but civilization which is the test of man's capacity for political rights'. APO provided a window into a community that would remain on the margins of South African's resistance movement for many years to come. The enthusiasm that had sustained the newspaper during the early years, gradually evaporated from 1911 onwards.


KORANTA EA BECOANA,

e Gatisioa mo Mafikeng Gangoe ka Tshipi.

MAFOKONYANA.

MAFOKO A NTOA.

3. *Koranta ea Becoana*.

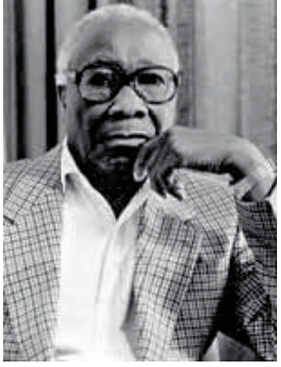
Black civil rights were heavily debated in British Parliament in the run-up to the establishment of the Union of South Africa in 1910. An anti-Union protest delegation consisting of newspaper editors Jabavu, Rubusana, Abdurahman and other black political leaders visited London in 1909. The impending Union of South Africa was characterized as 'The Great Betrayal'. Despite a major campaign by Blacks and Coloureds, the voter restrictions remained as in the pre-Union republics and colonies. Various discriminatory legislation were passed since 1892, restricting African voter rights (1892), the Pass Regulations Bill (1905), ownership of land (1913), the Urban Areas Act (1923), which introduced residential segregation in South Africa.

Abantu-Batho was the ANC's newspaper, established in 1912 by Pixley ka Isaka Seme. Published in English, Zulu, Xhosa and Sotho/Tswana. Although it was possibly the most influential of the black protest journals of this era, virtually no copies have survived.

1940's-1960's – From Protest to Resistance (African Nationalism)

African protest journals dominated the alternative press until the 1930s. Very few survived the Great Depression and political times of depression and repression in the mid 30s. The petitionary protest in South Africa was virtually bankrupt by the late 1930s. At the best of times black political publications had to struggle for survival in South Africa. The African nationalist press was relatively ineffective between the 1940s and 1960s.

The 1940s were critical years in the history of black politics in South Africa. Rapid industrial expansion and development during and after the war brought hundreds of thousands of Africans to the major cities in search of work and generated a highly politicized African working class. When the Nationalist Party won the 1948 election it was the beginning of the end for the protest journals as well as the political movements they represented. African nationalists remained divided. In future Africans would join together with non-Africans in a broader resistance movement to end apartheid in South Africa.



4. From left: Govan Mbeki; Ruth First.

Inkundla ya Bantu (1938-1951) under the leadership of Govan Mbeki and Jordan Ngubane, was the only major independent African nationalist newspaper left that played a significant role in the alternative press. A prime example of African liberal journalism.

As the potential market for African consumers expanded, and the state increased its efforts to institutionalise segregation and retribalize the Africans, white business interests moved in to take over surviving African publications. By 1946 between 250,000 and 500,000 Africans were reading newspapers. The vacuum created was filled by white entrepreneurs establishing a black commercial press with publications aimed at African readers.

Bantu World was launched in 1932 by Bertram Paver Published by Bantu Press, largely owned by the Argus, controlled by the mining industry. It was allowed a considerable degree of freedom to attack the government's racial policies. Bantu Press owned 10 African weekly newspapers by 1945. *Bantu World* in some ways was a trend setter in the shift from an elite to a mass audience. *Bantu World* and its imitators did not have to compete with the independent African political press, because there were virtually no publications of this kind left by the mid-1930s. *Bantu World* gave extensive coverage to black nationalist movements, even though it was never a protest organ. Black creative writing was offered an outlet in the newspaper, and virtually every member of the literary elite wrote for Bantu World at a time.

The Guardian (1937) represented a new non-African generation that mobilized the resistance movement. *The Guardian* was not the official mouthpiece of a political body. But the majority of staff and editors were probably members of the Communist Party of South Africa. *The Guardian* first of all was a mouthpiece for organized labour. It was cited by Albert Luthuli as 'the fighting mouthpiece of African aspirations'. A significant outlet for black grievances. Regarded as a platform for the whole liberation movement. The newspaper was banned several times. All its contributors were banned, jailed or forced into exile. Ruth First was the most investigative journalist during the *Guardian* era.

Digitisation at the NLSA

The NLSA was one of the founding members of DISA (Digital Innovation South Africa). DISA, which was established in 1997, aimed to implement digital technologies in libraries to enhance access to South African content of high socio-political interest (especially related to the Freedom Struggle). DISA became a centre of digitisation expertise in South Africa and provided training and support across South and Southern Africa. The NLSA contributed by way of digitising journals and other publications from its collections (e.g. Sechaba and Indian Opinion). As part of DISA Phase 2 the NLSA assisted with the scanning of archival collections that were published on microfilm.

5. NLSA equipment.

The NLSA transformed its photographic and microfilming facility to a fully operational digital services unit. Recently two large-format scanners were obtained with the purpose of scanning historical maps and newspapers. Other equipment include flatbed scanners, microfilm and film negative scanners, digital cameras and digital video technology. Staff are trained in the application of scanning technology and digital photography as well as multimedia and metadata.

Future plans for digitisation include the Library's collection of South African newspapers on microfilm. The microfilm collection consists of an estimated 7.5 million pages of newsprint. A selection of forty historical newspapers that represent the Black Press of South Africa of approximately 200 000 pages is considered.

References

Switzer, Les and Switzer, Donna. 1979. *The Black Press in South Africa and Lesotho: a descriptive bibliographic guide to African, Coloured and Indian newspapers, newsletters and magazines 1936-1976*. Boston, Mass.: G. K. Hall.

Switzer, Les et al. 1997. *South Africa's Alternative Press. Voices of Protest and Resistance, 1880-1960*. Cambridge: Cambridge University Press.

Training Program for Handling and Preservation of Microfilms and Photographs in Libraries and Archives Provided by the National Diet Library

by Shigehito Hisanaga, Preservation Division, National Diet Library, Japan

Purpose of Training

On June 30, 2011, the National Diet Library (NDL) held a training session, "Handling and preservation of microfilms and photographs in libraries and archives". More than 100 people from libraries, archives and museums attended. It included some participants from areas afflicted by the Great East Japan Earthquake. Microfilms are great media for long-term storage, but it is necessary to maintain an appropriate storage environment to get full performance out of them. Librarians, as users as well as managers of materials, are required to evaluate comprehensively what kind of measures are practical for long-term storage of microfilms. The purpose of this training session was to support librarians so that they can actively handle individual problems with view to long-term storage of microfilms by acquiring basic knowledge, understanding and lear ning conditions necessary for long-term storage, such as r ecommendations when producing microfilms, daily handling, regular maintenance and storage environment. In this course Mr. Nobuhiro Kuroki gave a lecture on the following items: 1) Deterioration cases and mechanism and condition for long-term storage of micr ofilms, and 2) Rescue project for photographs affected by the Great East Japan Earthquake. Mr. Nobuhiro Kuroki, member of the Qualification Program Committee of the Japan Image and Information Management Association is an expert on the structure and deterioration of microfilms.
In addition, Mr. Shigehito Hisanaga, Preservation Division, Acquisitions and Bibliography Department, NDL, made a presentation about the "Approach of the National Diet Library on the preservation of microfilm materials." The following is a brief overview.

Deterioration Cases and Mechanism and Condition for Long-Term Storage of Microfilms

Mr. Kuroki gave a lectur e on deterioration cases and mechanism of micr ofilms. He explained the basic knowledge of the structure and handling of microfilms and the mechanism of deterioration caused by them, and important conditions necessary for long-term storage with concrete examples of deterioration.

1. Structure and main component materials of microfilms

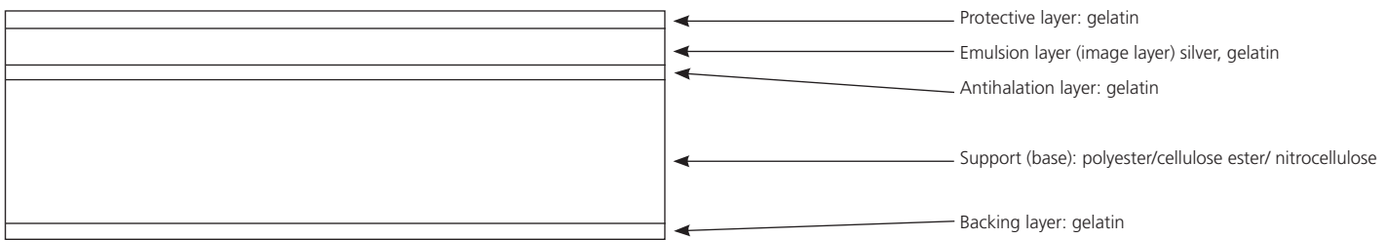
Microfilm is constructed as the following fi gure shows: fr om the top, protective layer, emulsion layer (image layer), antihalation layer, support (base) and backing layer.

If the right conditions ar e missing, each component material can undergo change and this leads to deterioration.

2. Main deterioration cases of microfilms

Microscopic blemishes and yellowing are cases of deterioration of developed silver. This occurs when oxidized developed silver transforms into silver ions which move within the emulsion layer and become minute yellow or reddish spots. This oxidization is caused by humidity and oxidized gas such as per oxide, ozone (O3), sulfite gas (SO2), hydrogen sulfide (H2S), nitrogen oxide (NO) and others in the air. Oxidized gas is considered to be derived from some sorts of building materials, plastics, gum, acid paper and emission gas as well as fixing defects or lack of water washing during the image processing procedure.
Crack, mold and sticking ar e cases of deterioration of gelatin. Humidity is the key element for each case. Cracks may occur when the r elative humidity becomes lower than 15%. Mold may occur when the relative humidity becomes higher than 50%. Sticking may occur when the r elative humidity becomes higher than 60%.
Deterioration of support, especially a cellulose ester base, is well known as vinegar syndrome. Ingredients of a cellulose ester base are cellulose triacetate itself and plasticizer for flexibility. Vinegar syndrome is caused by acetic acid pr oduced from hydrolyzing of the cellulose ester base, and its deterioration symptoms are seen going through the following process.

1. Hydrolysis occurs through incomplete drying of films during image processing procedure, and moisture from the humidity of the storage environment.
2. Acetic acid of the base splits off by hydrolysis.
3. Acetic acid promotes resolution and deterioration proceeds. Deterioration which reaches a certain level increases in speed.

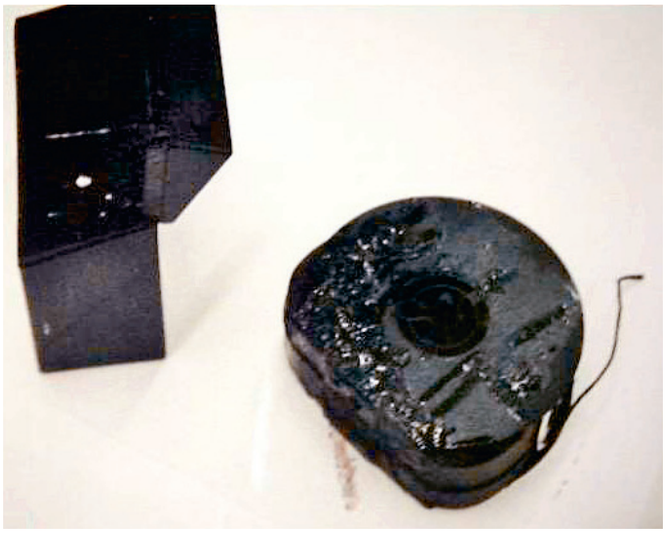


1. Structure and main component materials of microfilms.

2. Ruffling.

4. At the same time, plasticizer separates. Melted plasticizer crystallizes and when crystals get larger, they push the gelatin aside which leads to distortion and destruction of images and finally the images become unreadable. Films which lost plasticizer lose flexibility, and shrinkage and ruffling of the base occurs.

5. When deterioration proceeds, the plasticizer resolves. Resolution of the plasticizer produces acid materials, the pH level falls and deterioration proceeds faster. Gelatin resolves and it becomes a black mass exuding from the edge of films together with plasticizer and developed silver, or turns into black spots on the surface of the films. Sometimes even spools are dissolved.



3. Melted spool.

3. Storage conditions for microfilms

What is necessary to prevent the kinds of deterioration mentioned above and to store microfilms for a long time? Microfilms are recorded media with a long history and measures for long-term storage are defined by standard. For this reason, it is important to observe the standard for processing and storing microfilms.
Proper control of temperature and humidity is an essential condition to prevent deterioration of microfilms and for long-term storage. In the Japanese Industrial Standards (JIS), JIS Z 6009-1994 "Silver-gelatin type microfilms – processing and storage" is defined as follows:

4. Condition of relative humidity and temperature.

Storage conditions	Relative humidity (%)			Temperature (°C)
	Maximum	Minimum		Maximum
		Cellulose ester	Polyester	
Conditions for medium storage[1]	60	15	30	25
Conditions for permanent storage[2]	40	15	30	21

Ideally, temperature shall not be over 25° for a long time and it should be under 20°. The peak temperature for short time shall not exceed 32°.
Considering that:
1. The conditions of temperature and humidity shall be maintained 24 hours a day.
2. For permanent storage of cellulose ester base film and polyester base film in the same place, the recommended relative humidity is 30%.

To meet these conditions, measures such as setting up special storage, using cabinets with a function to control temperature and humidity, or using humidity control materials can be taken. In the International Standard, ISO 18911-2010, maximum temperature for the cellulose ester base is set extremely low.
For other conditions, it is necessary to keep down dust and gaseous impurities in the air and to pay attention to enclosures such as paper and plastics which are in contact with films.
Mr. Kuroki concluded that deterioration of microfilms is caused by three factors, materials factor, process factor and storage environment factor, and for long-term storage of microfilms, it is necessary to remove completely these factors of deterioration.

1. JIS Z 6009-1994 defines term of medium storage as 10 years or over.
2. JIS Z 6009-1994 defines term of permanent storage as permanent.

Rescue Project for Photographs Affected by the Great East Japan Earthquake

This training session was originally scheduled for March 18, 2011. Since the Great East Japan Earthquake occurred on March 11, the session was postponed to the end of June. Mr. Kuroki gave a lecture about the rescue project for photographs conducted by a major film manufacturer because there seemed to be urgent need for reconstruction from the earthquake and it would be of high interest among the participants of the training session. The following contents are as of the end of June when the training session was provided.

Many lives were lost in the tsunami caused by the Great East Japan Earthquake. Social infrastructure and buildings were swept up and reduced to piles of rubble. Cultural assets, books, documents, photographs and albums were also lost or damaged. In the case of cultural assets and documents, rescue projects have been continued by specialized agencies in local areas, neighboring prefectures and around the nation, and by volunteers technically supported by these agencies. The importance of rescuing photographs and albums has been recognized from early on, and the Self-Defense Forces and volunteers continue to retrieve photographs and albums out of the rubble and restore them based on the request from the government.

As there was a spate of inquiries to the customer service center of the film manufacturer after the earthquake about how to wash the muddy photograph, the technical department did research on deteriorated photographs contaminated by seawater and mud, and how to wash them off. They also conducted reproductive experiments using seawater and mud as well as figuring out the tools and disposables necessary for washing and checking the condition of the photographs by visiting volunteers who were actually collecting and washing damaged photographs in the disaster-afflicted area to offer more effective and comprehensible information. They posted and are continually updating on their website the information on how to wash the photographs without special tools or chemicals so that people in the afflicted area can clean their photographs. It was set up as a "Photo rescue project" and the following support activities are deployed.

One is a technical support activity of fering information on appropriate washing methods depending on the condition of the photo print. Details of the technical support activity are as follows: -Summarizing the washing procedure which is practicable in the afflicted area and providing it on a website and documents, and notifying people through TV and radio programs broadcast in the area (broadcast from April 23, 2011). Detailed information with videos is provided in several cases such as a water-stained and defaced photo, albums, negative films and inkjet-printed photographs, and several photographs stuck together, etc. For example, the following procedure is described: (1) soak them in water, (2) wash out the stains gently using a brush, (3) rinse them and (4) dry them in the shade, all these procedures with full attention to the safety. It is necessary to note that this procedure is a temporary and emergency measure and it could remove the image depending on the condition of the photo. This project also takes requests from local governments and volunteer organizations which are doing rescue activities for affected photos, offering its know-how of washing photos and giving advice.

The other is material support. Tools and disposables necessary for washing are provided to voluntary organizations, local governments and evacuation centers which are conducting photograph rescue work.

A contact point for photograph rescue projects was established to receive requests and deal with specific cases. It receives the inquiries via phone and the Internet and discusses the appropriate way to deal with individual cases. In addition, these support activities are done in cooperation with people in the afflicted areas, volunteer groups, photo studios, and various photo-related organizations in the area.

Over one hundred company staff members are cooperating with this project at weekends as volunteers, as well as 30 members continuing to offer technical support taking turns on site.

Approach of the National Diet Library on the Preservation of Microfilm Materials

At the end of the training session, Mr. Shigehito Hisanaga reported the approach of the National Diet Library on the preservation of microfilm materials.

1. Collection of the NDL

The NDL holds 8.84 million microfilms (as of the end of March 2011). There are 590,000 microfilms, 7.95 million microfiches and 300,000 micro prints. Certain parts of them were microfilmed by the NDL and others were acquired by legal deposit, purchase, donation and international exchange. The NDL makes both negatives and positives at once and basically, negatives are for preservation and positives are for use. It keeps the films for preservation in a dedicated microfilm storage. The temperature and relative humidity in this storage are maintained at 18°C and 25% in general with 24-hour air conditioning. The use of films for preservation is limited to recreating damaged positives, for digitization, etc. On the other hand, films for use are stored in the usual stacks along with paper materials such as books and periodicals. The temperature and relative humidity in the ordinary stacks are maintained at 22°C and 55%.

2. Project of preservation measures for microfilms

The NDL organized several preservation teams in 1983 by types of materials and reasons of damage and each team tackled a specific problem such as the physical damage caused by the increase of copying, the acid paper problem, etc. The NDL has preserved microfilms in the above-described environment, based on the results of discussion by the preservation team for microfilms.

After that, the NDL conducted two big projects for preserving microfilms. One was the recreating project of Japanese newspaper microfilms of 1990s. The other was the urgent countermeasures against deteriorated microfilm materials in fiscal years from 2004 to 2008.

2.1. Recreating project of Japanese newspaper microfilms of 1990s

Phenomenon such as acetic acid odor, deformation, tackiness, etc., was found in the Japanese newspaper microfilms (created between 1950s to 60s) in 1989. The NDL established an investigation committee including outside experts and investigated causes and discussed countermeasures. As a result, it was found that rapid deterioration, so-called "vinegar syndrome", caused by an inadequate environment for preservation of cellulose ester based films, happened. Therefore, countermeasures such as setting up the microfilm storage as noted above, recreating polyester base films, etc., were implemented. The target Japanese newspaper microfilms numbered over 30,000, so it was impossible to recreate them all at once. In order to delay the deterioration, we changed the metallic cases used until then to cases made of acid-free paper, and wound the films back to dissipate accumulated acetic acid.

2.2. Urgent countermeasures against deteriorated microfilm materials

When the NDL staff prepared to transfer materials to the Kansai-kan of the NDL when it opened in 2002, they found sticking microfiches of science technology reports (they were stored in the reading room where the temperature and humidity were not controlled) purchased from foreign countries. Staff of the NDL conducted a preliminary investigation to grasp the degree and dimension of deterioration. As the same phenomenon was confirmed in other microfilms, a five-year plan from fiscal 2004 to 2008 for urgent countermeasures against deterioration was decided upon. The target of this project was microfilms acquired by March 1992, which were possibly cellulose ester based. The NDL changed the way it created microfilm from cellulose ester base to polyester base in April 1992, and it was considered that there was no fear of degradation in polyester base film. The amount of the targeted microfilms was estimated at about 5,050,000. At the beginning, the NDL proposed to take the measures for all of them, but in the middle the number was reduced to about 1,800,000 of cellulose ester base microfilms for preservation.

The urgent countermeasures were implemented in two stages depending on the type of materials (books, periodicals, science and technology materials, rare books and old materials etc.). As a primary measure, replacing acid paper cases with acid-free paper cases, dissipation of acid gas by winding films back, and research of the deterioration level were conducted. As a secondary measure, restoration, recreation, and separation of films were conducted, based on the results of the primary measure.

After reporting the preservation measures of microfilms in the NDL, it was restated that three factors, namely good materials, proper handling and appropriate preservation are needed for long-term preservation of microfilms.

After Mr. Hisanaga's lecture, Mr. Kuroki explained the basic treatment of microfilms such as wearing gloves, and demonstrated the inspection and diffusion process by winding films back. The NDL also displayed deteriorated microfilms. The training session ended after an active question and answer session.

5. Mr. Kuroki demonstrated the inspection and diffusion process by winding films back, wearing gloves not so as to leave fingerprints on the microfilm, and a white mask on his face because of the acetic acid odor.

Events and Training

Announcements

78th IFLA General Conference and Assembly, PAC Open Session, 13 August 2012, Helsinki, Finland

At the occasion of the 78th IFLA General Conference which will take place in Helsinki, Finland, on the following topic: "Libraries Now! – Inspiring, Surprising, Empowering", the IFLA-PAC Core Activity will organize a two-hour session on **"Storage and Repositories: New Preservation and Access Strategies"** on August 13, 2012.

Programme

- "The Changing Face of Storage at the British Library"

by Deborah Novotny, Head of Collection Care, The British Library, London, UK

- "Automated Storage and Retrieval System: a Time-tested Innovation"

by Helen Heinrich, Chair of Technical Services, and Eric Willis, Library Systems Administrator, California State University, Northridge, USA

- "Moving to New Digital Storage: Migrating and Reloading Collections"

by Tanja de Boer, Head Collection Care, and Matthijs van Otegem, Head of Operations, Koninklijke Bibliotheek, The Hague, The Netherlands

- "The Creation and Upcoming Revision of *Archival and Special Collections Facilities: Guidelines for Archivists, Librarians, Architects, and Engineers, a National Standard of the Society of American Archivists.*"

by Diane Vogt-O'Connor, Chief of Conservation, Library of Congress, Washington, DC, USA

More information on the official website: http://conference.ifla.org/ifla78

International Council of Archives Congress, 20-24 August 2012, Brisbane, Australia

The National Archives of Australia will welcome the ICA Congress in August 2012, in the city of Brisbane on the east coast of Australia.
The aim is to provide an environment that stimulates debate, encourages the exchange of ideas and experience and acknowledges the challenges that face all archives in the 21st century.

The congress will examine our 'climate of change' through the themes:

- Sustainability: Archives recognising archival and information management challenges and working together on strategies to ensure access, preservation, security, and longevity of evidence and information.
- Trust: Archives supporting good governance and accountability, advocating ethical and professional processes, developing standards and gaining international acceptance.
- Identity: Archives helping the community to connect with their heritage, discover their individual stories and protect their rights; strengthening the value, impact and influence of archivists and information managers.

More information on the conference website: http://www.ica2012.com/

Contact: info@ica2012.com

International Conference "The Memory of the World in the Digital Age: Digitization and Preservation", 26-28 September 2012, Vancouver, British Columbia, Canada

UNESCO's Memory of the World Programme, in cooperation with the School of Library, Archival and Information Studies and with the Library of the University of British Columbia, and in partnership with IFLA, ICA, ICOM, WIPO, Google, Microsoft and others, is sponsoring a three-day conference concerning the preservation of documentary heritage. This Conference will provide a platform to showcase major initiatives that could lead to synergies both in research and implementation.

An open, dedicated space will be soon established on UNESCO's CI website for this event which will provide a restricted area to share documents.

The Conference will be opened by UNESCO's Director-General in the presence of the more than 500 participants who are expected to attend. English/French interpretation will be provided. Participation will be open to all those interested in heritage preservation: government decision-makers and policy planners, practitioners and professionals, as well as academics, legal specialists, information and digital technicians, representatives of the private sector, graduate students in the heritage disciplines, etc. Participants from developing countries are strongly encouraged to attend and some financial support may be provided to partially cover their expenses.

The ICOMOS Symposium, "Reducing Risks to Cultural Heritage from Natural and Human-Caused Disasters", 31 October 2012, Beijing, China

At the recent ICOMOS General Assembly in Paris, the interdisciplinary theme for the Scientific Council Triennial Action Plan for 2012-14 was discussed. Taking into consideration increasing risks to tangible and intangible cultural heritage due to various natural and human-caused factors, the themes for the scientific symposia for the next three Advisory Committee meetings will focus on risks resulting from natural and human-caused disasters (2012), globalization and uncontrolled development (2013), and loss of traditions and collective memory (2015). Consideration of risks also marks a shift from reactive to a preventive approach for conservation that seeks to put emphasis on risk reduction and preparedness.

The three themes will bring forward the underlying causes for risks to cultural heritage; tools and methodologies for their assessment; and policies, strategies and techniques for reducing potential threats to the future of cultural heritage aimed at protecting and managing our irreplaceable cultural resources for present and future generations.

Cultural heritage is exposed to numerous disasters resulting from natural hazards such as earthquakes, floods, cyclones, as increasingly human-induced hazards, such as arson, armed conflict and civil unrest. The great East Japan Tohuko Earthquake and Tsunami (2011); Thailand Floods (2011); Haiti, Chile and Christchurch earthquakes (2010); and recent civil unrests in Libya, Egypt, Yemen and Syria have caused serious damage to tangible and intangible attributes of cultural-heritage sites ranging from historic buildings, museums, historic settlements, as well as cultural landscapes.

Undoubtedly the frequency and intensity of some disasters has increased recently due to impact of Global Climate Change, as well as social, economic and political changes. Considering these challenges, The ICOMOS Symposium, Reducing Risks to Cultural Heritage from Natural and Human-Caused Disasters, aims to assess these risks and formulate policies, strategies and techniques for reducing risks to disasters, responding to emergencies and recovering from disasters. During the one-day symposium, position papers and case studies will be presented on the following themes:

1. Techniques and Strategies for Mitigating Risks to Cultural Heritage from Natural and Human-Caused Disasters
- How can we develop appropriate tech-

niques for mitigating risks to cultural heritage from earthquakes and floods, cyclones/hurricanes and fires by considering factors of safety, as well as values?
- What are traditional materials, skills and knowledge systems for disaster mitigation of cultural heritage, and how can we utilize them in present context?
- Which maintenance and monitoring strategies can be adopted for reducing risks to cultural heritage due to disasters?
- How can we enhance security of cultural-heritage sites to prevent risks of terrorism and theft?

2. Methodology and Tools for Undertaking Risk Assessment of Cultural Heritage
- What are various approaches and tools for assessing risks to cultural-heritage sites from natural and human-caused disasters?
- What are good practices in documentation, inventorying and mapping for recording and analyzing risks due to natural and human-caused factors?
- How can we communicate these risks to decision makers?

3. Protecting Cultural Heritage in Times of Conflict and Other Emergencies
- What kind of policies, techniques and strategies can be adopted for protecting cultural-heritage sites in the times of conflicts and other emergencies?
- How can we effectively use international legal instruments and coordinate with organizations such as Blue Shield?

4. Planning for Post-Disaster Recovery of Cultural Heritage
- How do we avoid hasty destruction of vulnerable materials and structures (earth, stone and wood) of architectural heritage located in disaster-prone areas?
- How do we undertake post-disaster damage assessment of cultural heritage?
- How can we develop monitoring and evaluation strategies for post-disaster interventions and reconstruction?
- How do we evaluate costs of post-disaster recovery and rehabilitation of cultural heritage?
- How do we engage various international and national stakeholders for post-disaster recovery of cultural heritage?
- How can intangible heritage be utilized effectively for post-disaster recovery and rehabilitation?

5. Awareness-Raising and Capacity-Building for Managing Disaster Risks to Cultural Heritage
- How do we engage communities for disaster-risk management of cultural-heritage sites?
- How do we build the capacity of craftsmen, professionals and decision makers for managing risks to cultural heritage from natural and human-caused factors?

In the first session, open to the general public, selected papers will be presented. Posters will be accepted as space and the blind peer-review process permits.
Part of the second session will be devoted to breakout groups for ISC members wherein each working group will be asked to reflect on specific topics and how they relates to their ISCs.
The breakout groups will return for a final plenary session to present each group's recommendations which will then be synthesized into formal recommendations to be distributed and discussed by the Advisory Committee and e-published for download on the ICOMOS website along with selected papers.

Call for Abstracts: "Reducing Risks to Heritage – International Meeting 2012", 28-30 November 2012, Cultural Heritage Agency of the Netherlands, Amersfoort, the Netherlands

Since 2005 ICCROM, Canadian Conservation Institute (CCI) and Cultural Heritage Agency of the Netherlands (RCE) (formerly ICN) have worked on developing and disseminating the risk management approach for cultural heritage and have organized six joint courses. The most recent course included, for the first time, a substantial distance-learning component to enable participants to apply the risk management approach in their own working and cultural context. The method and tools that have been developed for the course proved to be applicable for all types of heritage, ranging from single wall painting and large collections to historic buildings and archaeological sites.

At the end of this year the successful cooperation will be concluded with a meeting in which the experiences and gained knowledge from all these years will be shared among former course participants and with anyone else interested in risk management.

Background

Today, preventive conservation is widely recognized as a priority line of action. However, decision-makers are confronted with difficult choices in planning conservation strategies with limited resources. Should we put all our resources in a sophisticated environmental control system, or should we upgrade the fire control system instead? What exactly will happen to this collection of costumes and basketry if they remain exposed to this level of light? And in the long term, how will this damage compare to that caused by the increasing number of visitors? The risk management approach, which informs and guides decision makers in many other fields, offers a sound methodology to incorporate the most recent knowledge into current practice. It allows an integrated identification and analysis of all expected damages and losses to cultural property and a mitigation strategy to reduce these risks. It thus provides a useful tool for the design of more efficient conservation strategies.

Aim of the meeting and Program

The aim of the meeting is to synthesize and share knowledge and experience with former participants of the ICCROM-CCI-RCE international courses and exchange it with others who are interested in risk management. The meeting aims to consolidate and expand the risk network and explore new directions for the future. The three-day program will contain presentations, discussions, and social activities. The program will be made available by summer.

Location: Amersfoort

The meeting will take place at the Cultural Heritage Agency of the Netherlands in Amersfoort. The building which houses the Ministry of Culture's centre of expertise and support for preservation and management of moveable, built, archeological and landscape heritage in the Netherlands is a showpiece of modern architecture in a historic setting.

Registration and Abstract submission

You are invited to register for the meeting. There is a limited number of seats available and acceptance will be done on a first come, first serve basis. The registration fee of 300€ includes a book of extended abstracts, coffee, tea, lunch each day and dinner on Thursday. For former participants of the ICCROM-CCI-ICN/RCE 'Reducing Risks' courses the fee is 150 €. Details for payment will be made available later.

To register and submit 150 word abstracts in English please use the registration/abstract submission form (http://fd7.formdesk.com/archis/reducing_risks). Deadline for submission is 30 June 2012. The organizing partners may make a selection or give suggestions before final acceptance in the program. For questions about the abstracts you can contact Bart Ankersmit at: b.ankersmit@cultureelerfgoed.nl

Travel grants

Former participants of the 'Reducing Risks' courses can apply for an ICCROM grant. Contact iv@iccrom.org for details.

Travel and Accommodation

Participants will need to make their own

arrangements for travel, visas and accommodation. Amersfoort can easily be reached from the international airport Amsterdam Schiphol by train in 35 minutes (2nd class single ticket €10,00). Amersfoort offers a variety of hotels with rooms for two at €100-150 per night. A list of hotels will be made available soon.

Organising partners

Cultural Heritage Agency of the Netherlands (RCE): www.cultureelerfgoed.nl
International Centre for the study of the Preservation and Restoration of Cultural Property (ICCROM): www.iccrom.org
Canadian Conservation Institute (CCI-ICC): www.cci-icc.gc.ca

Contact:
Dr Bart Ankersmit
Senior Researcher | Research
Rijksdienst voor het Cultureel Erfgoed
Cultural Heritage Agency
Smallepad 5 | 3811 MG Amersfoort
Postbus 1600 | 3800 BP Amersfoort
Tel. 033 – 421 7421
Fax 033 – 421 7799

La biblioteca nacional de Venezuela ofrece residencias en preservación y conservación documental

La Biblioteca Nacional de la República Bolivariana de Venezuela y el Archivo General de la Nación, ofrecen estudios teórico-prácticos en preservación y conservación documental, dirigidos a los responsables de la preservación y conservación de fondos documentales, bajo la modalidad de residencias, con la finalidad de preservar y conservar la memoria documental de los pueblos.

Estas residencias tendrán dos ediciones anuales. La primera cohorte inició actividades el 10 de abril del presente año en la sede del Centro Nacional de Preservación Documental de la Biblioteca Nacional de Venezuela. Este Centro, pionero en el área de la preservación y la conservación, es reconocido como Centro Regional IFLA/PAC (Programa de Conservación y Preservación (PAC)) de la Federación Internacional de Instituciones y Asociaciones Bibliotecarias (IFLA) para América Latina y el Caribe desde 1988, y tiene entre sus principales objetivos fortalecer el conocimiento técnico en esta área a nivel regional.

Es por ello que estas residencias tienen una convocatoria abierta e internacional para la segunda cohorte que tendrá lugar durante el segundo semestre de 2012 con una duración de ocho semanas distribuidas en cuatro módulos: I. Principios Teóricos de la Preservación Documental, II. Preservación en sitio, III. Preservación de colecciones y IV. Conservación de colecciones. Es importante resaltar que la metodología utilizada en estas residencias supone el acompañamiento personalizado por parte de expertos en el área, lo cual facilita el proceso de aprendizaje.

Los participantes van rotando por los distintos módulos, a medida que van desarrollando el trabajo, con el acompañamiento del tutor asignado en cada una de las tareas propias de la preservación y la conservación documental; desde el diagnóstico hasta la restauración; considerando que cada documento debe tratarse de manera particular, de acuerdo con el daño que presente.

Los interesados, pueden visitar la página web de la Biblioteca Nacional de Venezuela, www.bnv.gob.ve en donde encontrarán un link con toda la información necesaria, o comunicarse a través de la dirección electrónica comite.academico@bnv.gob.ve o por los números telefónicos (58)(212) 5059341/5059018.

Report

IFLA International Newspaper Conference, 11-13 April 2012, Bibliothèque nationale de France, Paris, France

The Bibliothèque nationale de France, the IFLA Newspaper Section and the IFLA-PAC Core Activity organized in Paris the IFLA International Newspaper Conference 2012, in the BnF Grand Auditorium, on April 11-13, 2012. The topic was: "Newspaper Digitization and Preservation: New prospects, Stakeholders, Practices, Users and Business Models."

The Conference was aimed at assessing major ongoing mass digitization projects in Europe and throughout the world undertaken by libraries and archives but also by press groups, while dealing with the preservation strategies inherently linked to general digitization policies.

The Conference was opened by two keynote speakers, Emmanuel Hoog, President of Agence France-Presse, and Patrick Eveno, Media Historian, Pantheon-Sorbonne Paris 1 University, France.

The first day was dedicated to the very challenging task public institutions and newspaper groups have to face in order to store, preserve and provide access to their huge newspaper collections. Newspapers such as *Le Monde, Corriere della Sera* and *Ouest-France* presented their strategies in terms of digitization and all raised the same issues, including:
- the difficulty to manage the great number of editions.

1. Opening of the Conference: Jacqueline Sanson, Director General, BnF, Emmanuel Hoog, President, Agence France-Presse, Frederick Zarndt, Chair, IFLA Newspaper Section, and Patrick Eveno, Media Historian, Pantheon-Sorbonne Paris 1 University. © David Paul Carr/BnF

- the scanning source issue: microfilm or original paper pages, knowing that scanning from microfilm is faster and cheaper, but can raise some quality problems. Indeed, some microfilm pages may not fit the quality level expected, causing errors in article texts after the OCR process. Nevertheless, in most cases, rescanning from original paper page does not solve the quality issue, the paper itself being the problem.

Public institutions also presented their own digitization programmes, such as the BnF, and all agreed that, to fund the digitization of a growing and at-risk collection, a business model has to be found. One of the solutions proposed is the private/public sector partnership. The most representative case is the British Library's partnership with brightsolid which at their own cost and own risk invest to digitize millions of pages and establish a business model that will enable them to get a return on their investment.

The second day focused on Collecting and Access. Two different perspectives were presented: public institutions through legal deposit (the BnF presented its deposit of online Newspapers) and private aggregators such as Europresse.com, both facing new challenges due to the development of online press (particularly "pure players") and the increasing flow of information, which force to rethink the way of processing and storing. The other morning presentations dealt with recent developments of libraries in terms of access and new users practices, particularly crowdsourcing.

The afternoon session was completely dedicated to the archiving of Press Photographs, in public institutions such as the BnF and the National Library of Austria, and in agencies such as AFP and Getty Images. It is the same problem as for newspaper collection: the size of the collections to digitize (for instance, regarding the Getty Images Hulton Archive, only 0,6% of the collection is digitized), the importance of metadata, the issue of long-term preservation of digital files and storage capacities, to which must be added specific copyright issues, reinforced by the Internet revolution which leads to a growing use of images. Once again, a new business model has to be defined. For instance, the Austrian National Library chose to cooperate with the Austrian Press Agency (APA PictureDesk). The goal of this cooperation is to give access to the APA Press Photography to students and researchers at the Austrian National Library and to exploit internationally the rich historical photographic collections of the Austrian National Library via APA PictureDesk.

The third day was dedicated to the preservation of original and digital newspaper collections. Stress was put on the preparation of the paper originals before digitization and on the treatment of the physical collections, such as deacidification. The British Library showed that they chose to invest on the storage facilities with low-oxygen more than on mass deacidification treatment.

The last session focused on the long-term preservation of digitized and born-digital newspaper collections, which raises technical challenges, particularly about the volume to manage, the variability and heterogeneity of the data, formats and obsolescence issues, storage capacities and costs. The BnF presented its scalable preservation and archiving repository, SPAR.

The conference attendees also had the opportunity to visit a major exhibition dedicated to the history of Newspapers in France untitled "La Presse à la Une". A visit of the BnF Technical Centre in Bussy St Georges was also proposed on Friday afternoon, on registration. Moreover, during the whole conference, sponsors displayed their materials in the Auditorium foyer.

Christiane Baryla would like to thank the Newspaper Section, the Conference Scientific Committee, the BnF direction and staff who helped with the organization of the event and the Conference Sponsors: Zeutschel, CCS, Isako, Diadeis, I2S Digibook, Planman Technologies, Cedrom-Sni, Bookkeeper and Stouls.

The speakers' presentations are available on IFLA-PAC webpage at:
www.ifla.org/en/node/5932

2. Shalev Vayness, ISAKO, Claudio Albanese, IDM, Walter Colombo, Digitalizzazione Archivio *Corriere della Sera*, Jacek Brzezinski, *Ouest-France*, Sebastien Carganico, *Le Monde*, and Philippe Mezzasalma, Department Law, Economics, Politics, BnF. © David Paul Carr/BnF

PAC CORE ACTIVITY

USA and CANADA

LIBRARY OF CONGRESS
101 Independence Avenue, S. E.
Washington, D. C. 20540-4500 USA

Director: Mark SWEENEY
Tel: + 1 202 707 7423
Fax: + 1 202 707 3434
E-mail: mswe@loc.gov
http://marvel.loc.gov
http://www.loc.gov/index.html

PAC INTERNATIONAL FOCAL POINT AND REGIONAL CENTRE FOR WESTERN EUROPE, NORTH AFRICA AND MIDDLE EAST

BIBLIOTHÈQUE NATIONALE DE FRANCE
Quai François-Mauriac
75706 Paris cedex 13 - France

Director: Christiane BARYLA
Tel: + 33 (0) 1 53 79 59 70
Fax: + 33 (0) 1 53 79 59 80
E-mail: christiane.baryla@bnf.fr
http://www.ifla.org/en/pac

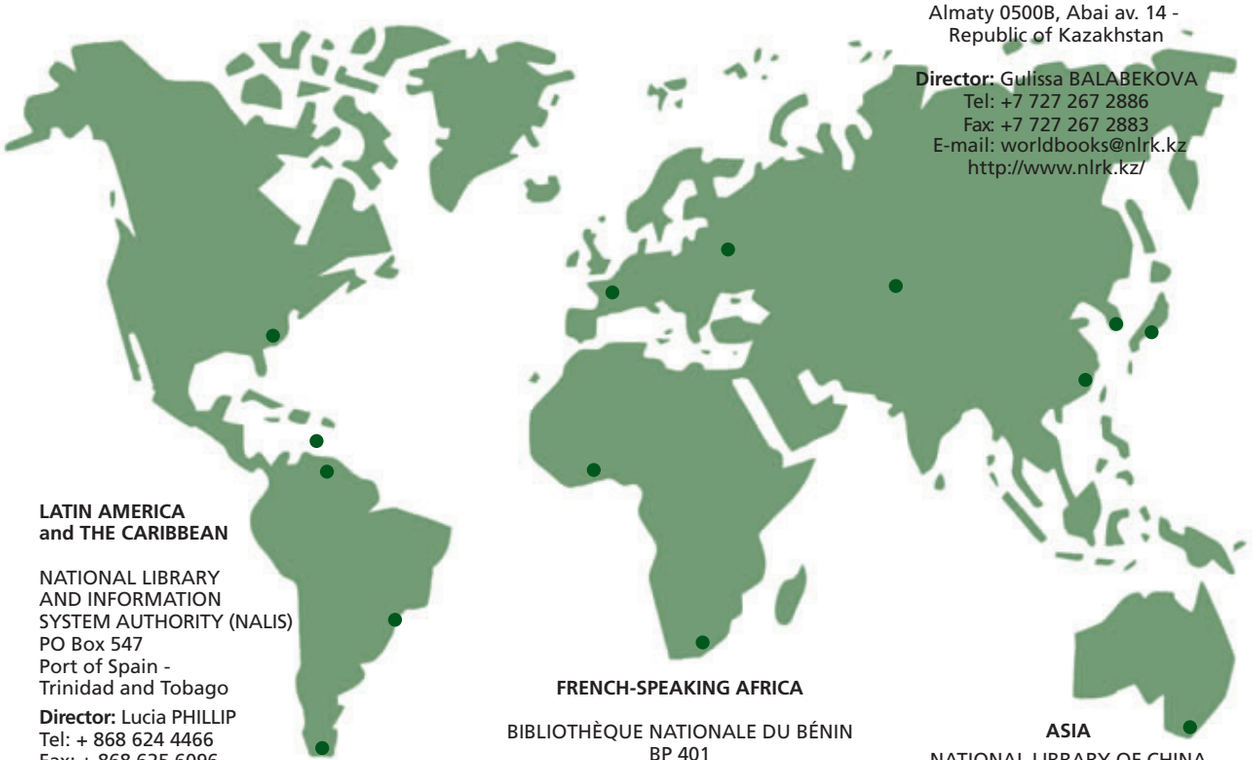
EASTERN EUROPE and THE CIS

LIBRARY FOR FOREIGN LITERATURE
Nikoloyamskaya str. 1
Moscow 109 189 - Russia

Director: Rosa SALNIKOVA
Tel: + 7 495 915 3696
Fax: + 7 495 915 3637
E-mail: rsalnikova@libfl.ru
http://www.libfl.ru/index-eng.shtml

NATIONAL LIBRARY
OF THE REPUBLIC OF KAZAKHSTAN
Almaty 0500B, Abai av. 14 -
Republic of Kazakhstan

Director: Gulissa BALABEKOVA
Tel: +7 727 267 2886
Fax: +7 727 267 2883
E-mail: worldbooks@nlrk.kz
http://www.nlrk.kz/



LATIN AMERICA and THE CARIBBEAN

NATIONAL LIBRARY
AND INFORMATION
SYSTEM AUTHORITY (NALIS)
PO Box 547
Port of Spain -
Trinidad and Tobago

Director: Lucia PHILLIP
Tel: + 868 624 4466
Fax: + 868 625 6096
E-mail: lphillip@nalis.gov.tt
www.nalis.gov.tt/

BIBLIOTECA NACIONAL
DE VENEZUELA
Apartado Postal 6525
Carmelitas Caracas 1010 - Venezuela

Director: Ramón SIFONTES
Tel: + 58 212 505 90 51
E-mail: ramon87s@hotmail.com
www.bnv.bib.ve/

FUNDAÇAO BIBLIOTECA NACIONAL DE BRASIL
Av. Rio Branco 219/39
20040-0008 Rio de Janeiro - RJ - Brasil

Director: Jayme SPINELLI
Tel: + 55 21 2220 1973
Fax: + 55 21 2544 8596
E-mail: jspinelli@bn.br
www.bn.br

BIBLIOTECA NACIONAL DE CHILE
Av. Libertador Bernardo O'higgins Nº 651
Santiago - Chile

Director: Maria Antonieta PALMA VARAS
Tel: + 56-2 360 52 39
Fax: + 56-2 638 04 61
E-mail: antonieta.palma@bndechile.cl
www.bibliotecanacional.cl/

FRENCH-SPEAKING AFRICA

BIBLIOTHÈQUE NATIONALE DU BÉNIN
BP 401
Porto Novo - Bénin

Director: Francis Marie-José ZOGO
Tel/Fax: + 229 20 22 25 85
E-mail: derosfr@yahoo.fr
www.bj.refer.org/benin_ct

SOUTHERN AFRICA

NATIONAL LIBRARY
OF SOUTH AFRICA
Private Bag X990
Pretoria - South Africa

Director: Douwe DRIJFHOUT
Tel: + 27 21 424 6320 ext 5642
Fax: + 27 21 423 3359
E-mail: douwe.drijfhout@nlsa.ac.za

OCEANIA and SOUTH EAST ASIA

NATIONAL LIBRARY
OF AUSTRALIA
Parkes Place
Canberra Act 2600 - Australia

Director: Pam GATENBY
Tel: + 61 2 6262 1672
Fax: + 61 2 6273 2545
E-mail: pgatenby@nla.gov.au
www.nla.gov.au/

ASIA

NATIONAL LIBRARY OF CHINA
33 Zhongguancun Nandajie
Beijing 100081 - China

Director: Zhang Zhiqing
Fax: + 86 10 6841 9271
E-mail: interco@nlc.gov.cn
http://www.nlc.gov.cn/en/services/iflapac_chinacenter

NATIONAL DIET LIBRARY
10-1, Nagatacho 1-chome,
Chiyoda-ku, Tokyo, 100-8924 - Japan

Director: Naoko KOBAYASHI
Tel: + 81 3 3581 2331
Fax: + 81 3 3592 0783
E-mail: pacasia@ndl.go.jp
www.ndl.go.jp/

NATIONAL LIBRARY
OF KOREA
KRILI/Preservation office
Banpo-Ro 664, Seocho-gu
Seoul 137-702 - Korea

Director: Guiwon LEE
Tel: + 82-02-535-4142
E-mail: leegw@mail.nl.go.kr