



IFLA International Newspaper Conference 2010

Digital Preservation and Access to News and Views

CONFERENCE PAPERS

IFLA International Newspaper Conference 2010



Editors
Ramesh C. Gaur, Frederick Zarndt
D. R. Gupta, Kavita Gaur



Indira Gandhi National Centre for the Arts, New Delhi
IFLA Newspaper Section



IFLA International Newspaper Conference 2010

**Digital Preservation and Access to
News and Views**

CONFERENCE PAPERS

(25th – 28th February, 2010)

Editors

Ramesh C. Gaur

Frederick Zarndt

D. R. Gupta

Kavita Gaur



Indira Gandhi National Centre for the Arts, New Delhi



IFLA Newspaper Section

Published by: *Indira Gandhi National Centre for the Arts*

IFLA Newspaper Section

February, 2010

New Delhi

Printed by: Om Laser Printers Delhi- 110009

Table of Contents

Physical Preservation of Newspaper Resources (Hardcopies Archives, Microfilm Archives) the practices at the Bibliothèque Nationale de France (The French National Library) <i>Else Delaunay</i>	1
Following the Crowd: User Engagement and the Australian Newspapers Service Experience <i>Cathy Pilgrim</i>	10
Newspaper Digitisation in South Africa <i>Patricia Liebetau</i>	18
Newspaper Digitisation : A Silver Bullet <i>Deborah Novotny</i>	25
Preservation and Archiving Solutions <i>Keith Rajecki</i>	33
24 by 7 Digital Access to Newspapers: The Singapore NLB's Experience <i>Ngian Lek Choh</i>	48
Improving Access to Digitized Historical Newspapers with Text Mining Coordinated Models and Formative User Interface Design <i>Robert B. Allen</i>	54
Newspapers as a Digital Resource National Infrastructure and Newspaper Publishers – in Finland <i>Majlis Bremer-Laamanen</i>	60
Connecting the Dots: How Researchers use their Library's News Resources <i>Debora Cheney</i>	68
Building A 24x7 Newspaper Digital Library and Archives (Case Study of DNA – Daily News & Analysis) <i>Anita Pujari</i>	90
Digital Preservation and Access to Print Media Resources: Experiences at the Times Archives and Knowledge Centre, India <i>R. Venkata Kesavan, Shankar Marathe, Salila Sreesan and Prof. Vijaya P. Rajhansa</i>	101
Newspaper Database Management at MICA's Knowledge Exchange & Information Centre <i>Shailesh R. Yagnik, Niraj R. Patel and Lavji N. Zala</i>	115
Online Newspaper Reading Habits among PhD Students and Faculty Members in Aligarh Muslim University <i>Naushad Ali P.M and Mohamed Musthafa K</i>	127
Open Access Model for Libraries and Newspapers: New Roles and Convergences <i>Ajit Pyati</i>	143

Newspaper Digital Libraries News-Clipping Services and Long Term Archiving Using Greenstone <i>M.G. Sreekumar, R. Biju, T. Sunitha, P. Sreejaya, Joshy Kuriakose and K.M. Sudheesh Kumar</i>	150
Media Archives-cum-National Reference Library on the North-East (India) <i>Prof Alaka Buragohain</i>	160
Online Bengali Newspapers: A Comparative Study <i>Ashis Biswas and Mayuri Das Biswas</i>	167
Access to Indian Online Newspapers: Problems and Prospects <i>Avijit Chakrabarti</i>	177
Access to Online Newspapers of India: An Overview <i>Dr. O.N. Chaubey and Dr. (Mrs.) R.Chandra</i>	188
Googling News and Views Online <i>Dr. D.R. Gupta</i>	195
Newspaper Section of the National Library Kolkata: Physical Preservation and Access to News <i>Gopa Ghosh, Nivedita Bhattacharya and Avijit Chakrabarti</i>	205
Building a Web Based Digital Repository of Images: Case Study of the Hindustan Times <i>Pratibha Kaushik</i>	215
Preservation of Newspapers in JNU Library: A Case Study <i>Dr. M. Natarajan and Payel Biswas</i>	226
Bengali Newspapers and Their Newspaper Preservation Techniques: An Overview <i>Pritam Gurey and Avijit Chakrabarti</i>	233
Growth and Development of Online Newspapers with Special Reference to India <i>Dr. Sumeer Gul and Tariq Ahmad Shah</i>	241
Concept Analysis of News for Mining Strategic Information <i>Sumit Goswami and V Senthil</i>	253
Training Programmes for Newspaper Library Professionals in India <i>Dr. G. Sivaprasad</i>	266
Archiving News in a University using Open Source Software <i>Surendran Cherukodan and Dr. Sheeja N.K.</i>	273
Press Clippings Information Service of the Jawaharlal Nehru University Library: A Case Study <i>Mahesh Chand</i>	280

Editors' note

We are pleased to present to you some of the papers of IFLA International Newspaper Conference 2010: Digital Preservation and Access to News and Views being organized jointly by Indira Gandhi National Centre for the Arts and IFLA Newspaper Section from 25th to 28th February, 2010. The volume includes a total of 29 papers both invited as well as contributed by more than 45 authors from 9 different countries. These papers cover various aspects such as physical preservation, digital preservation of both traditional as well as born digital, digitization, online newspapers, newspaper as digital resource and 24X7 digital access to newspapers. Also explored are various issues concerning digital archiving of newspapers and the reading habits of users in the digital era.

The papers in this volume are presented with a view to triggering further discussions during the Conference. Later this volume may also help the participants as a reference tool in implementing some of the ideas discussed during the Conference. This is the first time that such an International Conference is being organized in India. Therefore, we believe the newspaper- libraries and archives in India would be able to derive much value out of the contents of this volume.

We have tried to include almost all the papers being presented in the Conference. However, due to the paucity of time some of the papers have not been included. We understand that a complete volume will be published by IFLA Newspaper Section as proceedings of the Conference in due course of time and some of the papers which have not been included here may be part of the proceedings.

The deadline for receiving of full text papers was 7th February, 2010. However, many of the papers have been received till 17th February, 2010. So, we were having few days to edit these papers. Despite our best efforts some errors may have crept in, so due apologies to Authors for any such errors.

We are grateful to all the speakers at the Conference for providing valuable inputs for this volume. Our sincere thanks to the Printer of this volume for timely printing and special thanks to Mr. R.S.Kaushik for his help. The typing assistance received from Mrs. Kiran Kapoor and Mrs.

Sunita Arora IGNCA is duly acknowledged. Last but not the least we are grateful to the Members of the Publication Committee of the above Conference, namely, Ramesh C. Gaur, Frederick Zarndt, Par Nilsson, Edmund King and Hartmut Walravens for their contributions in screening and review of proposals received for the above Conference. Without their efforts this volume may not have been possible.

Editors

From the IFLA Newspapers Section Chair

This year's IFLA Newspapers Section conference, on the theme "Digital Preservation and Access to News and Views", takes a progressive look at news and newspapers. In keeping with the digital focus of Stockholm's August 2009 conference on "The Present becomes the past: Harvesting, archiving, presenting today's digitally produced newspapers", the 2010 New Delhi conference also focuses as much on collection, preservation, and access of digital news and newspapers as it does on these same activities for traditional print newspapers.

The reason for this is obvious: News is now created and produced digitally, even for print newspapers. Almost without exception, mid-sized or larger newspaper publishers distribute news via the Internet concurrently with their print issues. And although news continues to be distributed in print, some newspaper publishers such as the Christian Science Monitor have stopped printing daily newspapers; other publishers have eliminated the print issue altogether, and still other publishers are experimenting with distribution via devices such as Amazon's Kindle.

The Newspapers Section has in past been 'concerned with all issues relating to newspapers in libraries and archives, including acquisition and collection development, intellectual and physical access, storage and handling, preservation of newspapers and their contents'. Now it faces new tasks such as collection (harvesting) and preservation of online news and born digital newspapers and providing access to it as well as historical print newspapers. These tasks are not trivial.

With the daily or weekly print newspapers, one knows what must be collected and preserved -- it has a physical instantiation -- even if one does not always have the resources to do so. However content in born digital news and newspapers and especially in online news can be changed and updated so rapidly that it is not possible to collect or preserve it with traditional methods and technologies. Furthermore digital news is fragile: Bit rot corrupts files. Digital photos and digital content are only a hard disk crash or a virus infestation away from destruction.

This present era of transition from traditional newspapers to digitally produced news offers both opportunities and challenges. New thinking, new methodologies, and new technologies are needed to cope with the challenges of digital news production and distribution. But with the seminal efforts already made by cultural heritage organisations around the world, with the further help of the librarians, researchers, practitioners, and archivists whose papers are collected in this volume, and finally with the further efforts of you the reader and conference attendee, these necessary innovations shall be discovered.

Frederick Zarndt, Chair
IFLA Newspapers Section
Coronado CA USA
February 2010

From the Conference Director

It is indeed an honour to extend a hearty welcome to all the participants, speakers and eminent guests at IFLA International Conference 2010, being jointly organized by Indira Gandhi National Centre for the Arts and IFLA Newspaper Section from 25th to 28th February, 2010. This is for the first time that such an International Conference is being organized in India. When I was contacted by Mr. Hartmut Walravens former Chair, IFLA Newspaper Section for organizing this Conference at IGNCA, I was not sure of accepting this proposal, as neither did I have any background in Newspaper Libraries nor does IGNCA have any programme for preservation of newspapers. Later, I accepted it because I found this an opportunity to create awareness about preservation of and access to newspaper-based information services in India.

Initial response to the Conference information circulated by me on various listings was reasonably encouraging. But the kind of response I received in terms of participation, papers and sponsorship is amazing. Let me tell you that major cost expenditure with regard to this Conference has been taken care of with the help of sponsorships from the industry including registration fee of the participants and speakers. The delegates to the Conference are from libraries of more than 15 countries such as Library of Congress, British Library, National Library of Australia, National Library Board, Singapore, National Library of France and Denmark. Coming back to our own country, along with the major reputed newspapers i.e. The Times of India, The Hindustan Times, The Hindu, Indian Express, DNA, Deccan Chronicle, several regional newspaper libraries such as those of Matrabhumi, Eenadu, Rajasthan Patrika are also participating in this Conference. We have also received registrations from some media persons in this conference. Apart from Newspaper Libraries and Archives, we have representation from National Library of India and some important Universities of India. I am sure with such a broad-based participation, the Conference will be able to achieve the objectives set out for this event of international significance.

This Conference has been organized with the efforts and cooperation of many individuals and institutions. It may not be possible for me to name everyone. However I would like to put on record the critical support received from the following individuals and institutions. Let me begin

with IFLA Newspaper Section particularly Mr. Frederick Zarndt for having faith in me for organizing of this Conference. I am grateful to all the authorities of IGNC A, Officers and staff for their constant support and encouragement in the organization of this Conference. I sincerely thank all sponsors for their financial support for this Conference. I would like to specially put on record the support rendered by Mr. Vishal Salgotra and Mr. Bharat Joshi of Planman Technologies in the organization of this Conference. I am grateful to all my colleagues from media libraries, namely, Mr. Dharam Vir, Mr. R. Venkata Kesavan, Mrs. Pratibha Kaushik, Ms. Vijaylakshmi, Mr. Pranav Priyadarshni, Ms. Anita Pujari, Mr. K. Rajindrababu for their support in making this Conference a real success. Excellent efforts by all officers and staff of Kala Nidhi Division of IGNC A are duly appreciated and acknowledged. I also appreciate the support from the Ministry of Culture, Ministry of Home Affairs and External Affairs in granting permissions and clearances for foreign delegates. I would also like to express my gratitude to all the participants, speakers, Chairpersons of various technical sessions and Rapporteurs for agreeing to contribute to this Conference.

For the last few months I have been working for 18 hours a day despite a great personal loss (of my father). I had to devote a large amount of time which actually belongs to my family. My heartfelt thanks to my wife, Mrs. Kavita Gaur and to my daughters Ritu Gaur and Kanika Gaur for allowing me to use their time for the organization of this Conference. Without their support all this could not have been possible. Last but not the least, I am grateful to all those whose names I may not have included here but their efforts were nonetheless there in the successful organization of this Conference.

Ramesh C. Gaur
Conference Director &
Head – Kala Nidhi Division, IGNC A
IFLA International Newspaper Conference, 2010
February, 2010

PHYSICAL PRESERVATION OF NEWSPAPER RESOURCES (HARDCOPIES ARCHIVES, MICROFILM ARCHIVES)

The practices at the Bibliothèque Nationale de France (The French National Library)

Else Delaunay

INTRODUCTION

In a digital era why should we keep hardcopies which are digitised and/or microfilmed and why should we preserve the original negatives when a given title has been digitised? New attitudes towards these questions seem to occur to day, all the more so, as electronic archiving still is problematic. Preservation of hardcopies and master negatives in optimum conditions is then of prime necessity. For a long time yet the hardcopy will be considered as a last resort.

It is necessary to distinguish between national libraries in charge of the national heritage, mostly received by legal deposit, and other big libraries such as university or specialised libraries which may be free to decide on the destiny of the hardcopies that have been microfilmed or digitised.

Storage facilities, in particularly in big towns, are very costly. Dynamic management of the stacks of a storage repository seeks to find new space for future increase of the holdings. If all hardcopies are to be preserved, an off-site silo with optimum preservation conditions may be the right solution. If not, the weeding of the collections may give the best and easiest result.

GENERAL OVERVIEW OF NEWSPAPER PRESERVATION METHODS

In general newspapers have no initial protection so they need to get some kind of packing to be preserved from wear and tear, light, dust... and also to facilitate the filing of the items on the shelves.

MOST IMPORTANT AND THE MOST FREQUENT DAMAGE FACTORS OF NEWSPAPERS ARE:

- wear and tear
- self-destruction due to the acidity in machine made paper
- large size item on low quality paper printed with poor quality ink
- poor storage conditions
- air conditions, pollution, etc.

Protection of the items

There are various protections of newspapers among which the alf-industrial binding is valuable but a bit too expensive (around 40 €taxes included)

For the boxing of newspapers, most libraries generally use preservation boxes in acid free non pigmented cardboard, of the same size as the newspaper. The long experience of boxing shows the quality of this protection and its price is quite low (around 10 €taxes included). The polypropylene boxes have proved not to be strong enough for newspapers which continue to be used by the readers. They are not expensive (around 2 €taxes included) but do break quite easily after some use.

The Maintenance of the Holdings

Maintenance is essential for the good preservation of the items and their protection. The staff in the stacks need to be particularly involved, as they are in daily contact with the holdings. They can take responsibility for various checks:

- supervision of the stacks and of the items sent to the reading room (e.g. progress of wear and tear, damaged items, etc.); detecting all anomalies or deteriorations;
- regular checking of thermo-hygrometers if not already checked by the Preservation Office;
- carrying out of necessary moves of the holdings, partial stocktaking, dust removal.

Parameters for optimum storage facilities are

- Cleaning and any necessary disinfection of the premises; airing
- Room temperature between +16° and +18° C (in warm countries between +20° and +22° C); relative humidity: 50 to 55 %
- Light level: not more than 50 lux
- Air conditioning, checking of pollution and dust
- Permanent checking of risks: (fire, flood, infesting, robbery ...)

Equipment of the stacks

Traditional or compact shelving; boxes or bound volumes of large size newspapers should be stored flat on the shelves with 3 boxes or volumes per shelf; transport of the items can be manual or automatic.

Collection building: estimates of the growth of holdings are particularly tricky for newspapers.

RESTORATION OF BRITTLE NEWSPAPER FILES**Deacidifying prior to restoration**

- **Manual deacidifying:** each sheet is immersed in a borax water solution, then dried in the open air or on special crates.

- **Mass deacidifying:** the **Wei T'o** process was developed in Canada. The base product is the methylmagnesium carbonate.
- The **Bookkeeper** process developed in the USA is used by the Library of Congress. In France the Bookkeeper process has been used to deacidify running regional dailies (1945-1970). At the end of the process an alkaline buffer is deposited directly on the fibres. For newspapers in a reasonable physical condition this process may give them an additional lifespan of 30 or 40 years.
- **Mass deacidifying in BNF's (Bibliothèque Nationale de France) technical Centre** in Sablé-sur-Sarthe (230 km West of Paris): the method applied to the items is inspired by the Wei T'o process. Around 600 volumes (in-octavo) are neutralized in a single process. The volumes are placed in baskets inside a sterilizer. An alkaline buffer is brought into the paper simultaneously. But this method cannot be applied to large size items such as newspapers.

SEMI-INDUSTRIAL RESTORATION

Thermosizing/laminating

As far as restoration of newspapers is concerned, the process generally will be manual deacidifying (in a borax water solution) followed by laminating (covering of each sheet recto-verso with ABF, a sheet of polyamide, and size, by means of a thermosizing machine). At the BNF, most heavy restoration of newspapers is carried out in BNF's technical Centre in Sablé. At the new technical Centre in Bussy Saint-Georges (30 km East of Paris) new developments have been tested concerning mass deacidifying and paper strengthening in one and the same process (use of supercritical CO₂ ; paper splitting). Considering newspapers for this type of treatment is particularly tricky because of the size of the item and the poor quality of paper and ink. The laminating process is quite expensive: some 300 sheets cost around 1100 €. Some 1000 sheets per year are restored according to this process.

Leafcasting

It is another way to restore brittle sheets. It is used at the technical Centre in Bussy Saint-Georges. After manual deacidifying, the sheets are immersed in a water solution of paper pulp mixed with tylose in a special machine so as to be reinforced by leafcasting. The sheets are dried hanging on strings in the open air. It is a costly process (only some 100 sheets yearly are restored in this way at the BNF).

Deacidifying and paper strengthening in a single process: Splitting

Paper splitting has been used for many years to restore old damaged books. The Deutsche Bibliothek in Leipzig in Germany together with the Becker Company has developed an industrial method of paper splitting for large size items using a huge machine which is very expensive. After deacidifying and splitting of a newspaper sheet a thin Japanese support tissue sheet is inserted between both

parts of the newspaper sheet so as to strengthen it. The physical condition of the item must not be too brittle. It does not seem to be a real mass process yet.

Due to lack of funding the BNF does not anymore restore newspaper sheets; if not needed for reproduction using microfilming or digitization process.

SAVING BY CONVERSION TO ANOTHER STORAGE MEDIUM: MICROFILM

The transfer of the hardcopy to a substitute, the microfilm, seeks to preserve the hardcopy from wear and tear. From the beginning, some 50 years ago, the process was applied to much used holdings whose physical condition was precarious. But items which are difficult to use, such as regional daily newspapers with many local editions, or, some pamphlet volumes, or artificial collections - these are microfilmed also.

Newspaper Preservation by Microfilming²

What exactly does preservation microfilming mean? It refers to documents which are generally in danger of reaching such a bad physical condition so that their transfer to another medium is necessary to save them from total destruction. Thus preservation microfilm should be an accurate image of the original document. A preservation microfilm is a substitute for the original which may be stored far from the library and far from wear and tear; in some cases the library may even decide to dispose of the original once it has been microfilmed. Preservation microfilming should use a 35 mm unperforated film with silver gelatine emulsion on a polyester base³, type black and white, with 70 cm leader and end slip, start and end technical targets and bibliographical targets on the film; 30 m universal reels in acid-free and lignin-free cardboard or polycarbonate boxes.

For newspapers of smaller size, it is possible to use a 16 mm silverhalide polyester film but the strength and the quality of legibility and of photocopying will not be as good, as for 35mm microfilm. In general, microfiche should not be used for the copying of newspapers. Preservation microfilms should always be made in three generations: master negative (archival master copy); internegative (working negative for duplications); positive copy, which is a service film for the users.

Completing of the files to microfilm

When a complete file is not available, missing issues should be sought for through interlibrary lending (often a difficult task), or by photocopying. In some cases it may be necessary to use some issues from another chronological edition in order to obtain a complete file. If so, these issues should be clearly announced in the bibliographic target. Because of the high cost of microfilming, it is important to make films which are as comprehensive as possible. Later completing of the run of a newspaper on microfilm (by splicing) is always a long and expensive process.

Reading equipment

The BNF has a great number of reading machines -- Gideon 1000 made by par Bell & Howell: costing about 4500 €(taxes included), and a few reader-printers to obtain photocopies from microfilms.

Storage requirements for preservation microfilms

Clean dust-free stacks especially fitted for storage of microfilm reels:

- master negative : low room temperature (14°-16° C; Relative Humidity 40 %)4, preferably in remote stacks. The master negative should never be stored in the same place as the working negative and the hardcopy when the latter is preserved by the library. Regular inspection of master negatives every three or five years is required;
- working negative: low room temperature (16°-18° C, Relative Humidity 50 %), possibly in remote stacks if not used regularly for duplication for commercial or library purposes;
- service film (positive duplication): room temperature of 18°-20° C, Relative Humidity 55-60 %, preferably in stacks close to the general reading room or to the special reading room for microforms.

Storage furniture

The reels of microfilm should be stored in cupboards, drawers or on ordinary shelves (e.g. in special boxes each one containing 6 reel boxes) allowing conditioned air to circulate. If the master negatives are fitted in hermetically sealed boxes, they may be stored on ordinary shelves. All storage furniture should be non combustible.

Microfilming in France

Until now there has not been any real national microfilming programme in France. The BNF has accomplished a good deal of preservation microfilming of its historical newspaper files but there is still much to do. In the last ten years, digitisation has been viewed as the new surrogate. Around 30 national historical newspapers, already microfilmed, are now being digitised. Digitisation offers immediate access to the items, but, owing to uncertainty about the durability of digital files, they do not supersede the microfilm as a long-term preservation means. However, it becomes hard to get funding for preservation microfilms. There is so much to be done in both sectors and funding is limited. Shared preservation of hardcopies and microfilm master negatives may be the best way to assure the right preservation of national heritage as the BNF alone cannot assure this anymore.

BNF's yearly microfilming programme of running newspaper files covers 11 national dailies, 20 French regional dailies with local editions of which 8 are filmed on behalf of the BNF and 12 acquired directly from the newspapers' microfilming agencies, the whole amounting to some 5500 reels. The microfilms of 19 titles of regional dailies with multi-editions that the BNF does not acquire anymore should be preserved and microfilmed by the regional legal deposit

libraries, the BDLI (Bibliothèques du Dépôt legal Imprimeurs), according to the recent amendments to the law on legal deposit.

In 2009, BNF's microfilming programme of newspaper backfiles included some 290 titles, meaning some 1.100000 images or 2200 reels. The programme focused on newsprint from ancient French colonies, on political newsprint and on backfiles of local French newspapers. The general total of the Programme was 145000 € which, of course, is not enough according to the preservation needs and the intensive use of the collections. In 2010 funding will be cut down.

Some foreign daily newspapers are required on microfilms by subscription for a yearly total of 115000 €

As a general rule, hardcopy originals which are reproduced on microfilm should not be sent to the reading-rooms anymore. Use of hardcopies is to be considered as a last resort in case the microfilmed or digitised copy is deteriorated or lost.

Some national plans of newspaper microfilming

Great efforts to microfilm all national historical newspaper files have been made in different countries: USNP (United States Newspaper Program)⁵, a national programme of cataloguing and microfilming of all American newspapers from the very beginning (launched in 1994 and almost finished in 2004); Newsplan⁶, a microfilming programme of all British historical newspapers since the 17th century was launched in 1984 and became effective from 1991 on. The four national libraries in the UK are in charge of the carrying out of the programme in the region, or the regions, under their responsibility. In the Netherlands the Metamorfoze⁷ programme was launched in 1997. It does include newspapers. It is not only a microfilm programme, it also concerns the preservation of the originals. The Nordic countries have microfilmed their national historical newspaper files from 1640 to 1900. The microfilms are now used for digitisation within the project TIDEN⁸ that has been open to the public since October 2001. The archives of the master negatives, of course, are safe.

Mass production techniques for the conversion of microfilm to digital images are used when the microfilms available are of high quality: according to Finnish experience, high-contrast films with a limited dynamic range are often a good feature when you compare the scanning result of the originals and the microfilms of grey text on brittle and brownish paper. Smaller shades, spots or wrinkles in the paper are also eliminated. The text is clearly extinguishable from the base. It favours full text search possibilities of the newspapers using OCR software on bi-tonal images. On the other hand, digitising from photographic material may not be as good as digitising in greyscale from the original... Microfilming for digitisation, full text search for future automatic scanning process: the use of International and National Standards for Microfilming and IFLA Guidelines and its supplement can be considered.

A hybrid approach: Digital Imaging and Preservation Microfilm

This technique could also be a mass production technique for conversion of digital images to microfilms in a single process.

The case of Ouest France, the first daily newspaper in France in terms of print run (with 42 local editions and 9 Sunday editions): this newspaper company produces microfilms and digital files using a hybrid solution: from born digital data to microfilm¹⁰. It is the only French newspaper which itself microfilmed all issues since 1899 when its predecessor Ouest Éclair was founded, until 2009, a total of around 9 million pages. From 2010 on Ouest France will go over to digitisation only.

In 2002, it was decided to microfilm directly its DTP layout files yielded for printing: instead of two captures and two processings, one continuous graphic chain (from the publisher's site to the provider's site, in this case ACRPP (the Association for Conservation and Photographic Reproduction of Newsprint), located in Croissy-Beaubourg, East of Paris).

Advantages of the microfilm

- The life-span of the black and white microfilm, well preserved and regularly checked, is of some 200 to 400 years or more
- A microfilm of high quality (with high contrast and high resolution) fits well for digitisation; digitisation from microfilm seems to be of even better quality and also less expensive (cf. Finnish experience).

Inconvenience of the microfilm

- The film is always in black and white so coloured illustrations in the hardcopy are reproduced only in black and white. Here the microfilm is not a real substitute to the original. Coloured microfilm can be used but it is three or four times more expensive than the black and white film, and its lifespan is only around 30 or 40 years. It is difficult to stabilize the colours.

SETTING UP OF A NEWSPAPER PRESERVATION PLAN

A preservation plan may include one or several programmes. When a programme has been decided and launched after detailed cost estimates, it is essential to follow it closely in order to avoid all supplementary expenses. In other words, it is necessary to meet the requirements included in the set of detailed specifications.

Cost of restoration

Preservation by conversion to microfilm supersedes more and more preservation by restoration as far as newspapers are concerned. The price of manual restoration is indeed very high. When needed, partly restoration by laminating is used for newspapers to microfilm or to digitise.

Cost of conversion to microfilm

All reproduction has a price. When it is easy to calculate the price of an image in three generations proposed by the microfilm provider, it is much more difficult to estimate the real price including all the various tasks before and after the

filming (ex. selection and setting apart of the items, collating page by page, preparing of bibliographic and/or technical targets, making a set of detailed specifications, completing of lacking issues, restoration of some sheets, various checkings after filming, boxing, etc.). All these tasks are important in order to get quality films but the price to pay is high. It will probably increase the image price by 200 or 300 % or even more.

SHARED PRESERVATION: SOME EXAMPLES IN FRANCE

For some kinds of collections, however, another possibility can be shared: dynamic management and preservation of the collections within a network of library partners. In France a documentary map was elaborated so as to know which subjects are covered by each library. A single repository, the CTLeS (Centre technique du livre de l'enseignement supérieur), was set up some ten years ago in Bussy-Saint-Georges (East of Paris) for the university and research libraries in Paris and the region Ile-de-France to share the preservation of their holdings. An agreement has been signed between the Centre and each library partner. After each weeding of the collections, a library will send its items to the centralised repository. One or two copies of the item should be kept. The repository is responsible for the best preservation, including restoration and/or microfilming, and it provides copies of the items on request. It also provides the library with technical advices.

Shared preservation may also be the result of a common decision by several libraries within a region or a special subject field. As an example the Shared Preservation Plan of Periodicals kept by 25 medical libraries within the region Ile-de-France, around 7000 periodicals. The Plan aims to: reduce preservation costs; create more space for the increase of holdings; achieve the best rational and long-term preservation for a limited number of holdings of each title; improve access to periodical holdings for document supply; promote the transfer to all electronic environment for the institutions which are not preservation poles. Shared preservation from an international point of view may also be considered by national libraries (ex. within the European Union or the Francophonie, a network of national libraries from French-speaking countries)

CONCLUSION

To conclude, I should say that, considering the high cost of restoration and reproduction of original hardcopy newspapers, the preservation of the national heritage of newspapers requires a close cooperation between libraries and archives. It will be important to decide firstly what hardcopy newspaper originals should be preserved, and secondly which originals may only have their information preserved in a surrogate format such as a microfilm or digital copy. A national shared preservation plan including regional and local programmes for restoration and reproduction should be organised and carried out. To share the cost and the work seems to be the way to preserve and save as much as possible of the newspapers within a country. National libraries alone cannot assume this task anymore.

SOME INTERNATIONAL STANDARDS

ISO 4087 : Micrographics – Microfilming of Newspapers for Archival Purposes on 35 mm Microfilm. 2nd edition, 1991

ISO 10602: Photography – Processed Silver Gelatin Type black and white Film. 1993

ISO International Symbols 7000/...

ISO 5466: Photography – Processed Safety Photographic Film. Storage Practices. 1988

1. International Newspaper Conference IGNCA – New Delhi – 25-28 February 2010
2. IFLA Round Table on Newspapers/IFLA Section on Serial Publications., 1996. Guidelines for Newspaper Microfilming. *IFLA Professional Reports, N° 49.* – The Hague, 1996, and *Supplement 2001: Guidance for Preservation Microfilming of Newspapers for Digitisation.*
3. ISO 10602:1993
4. As reference n° 2 above
5. <http://www.neh.gov/projects/usnp.html>
6. <http://www.newsplan.co.uk>
7. <http://www.metamorfoze.nl/en/programma/index.html>
8. <http://tiden.kb.se/>
9. As reference n° 2 above
10. Chapman, S., P. Conway & A.R. Kenney. Digital Imaging and Preservation Microfilm: The Future of a hybrid approach for the Preservation of Brittle Books. Washington D.C. CLIR, 1999
<http://www.clir.org/pubs/archives/hybridintro.html>
11. contact@acrpp.fr

FOLLOWING THE CROWD: USER ENGAGEMENT AND THE AUSTRALIAN NEWSPAPERS SERVICE EXPERIENCE

Cathy Pilgrim

ABSTRACT

The National Library of Australia has a well-established history of collecting and preserving Australia's documentary heritage in its many forms. The Library also has another role, one that is increasingly important in a world of rapidly changing technology – we must ensure that the wonderful collections can be accessed by anyone for research and study, but also for the pleasure of discovering stories and images of Australian history and society, irrespective of geographic location.

How the Library shares this material with the Australian and wider international community is changing rapidly as digital and online technologies advance. Today users can search the Library's collections and services online, view digitised collection items and related material and use the Library as a gateway to the collections of over 800 other Australian libraries. The Library's primary goal is to become more visible in the online information-seeking world in order to reach new audiences and expand the community's awareness of our collections and their significance.

This paper will explore the innovative uses of new technologies the Library has made in recent years, to improve access to Australian newspaper content. With the Australian Newspapers Digitisation Program (ANDP), the Library is not only supporting users to locate relevant material online, but is also enabling users to enrich and enhance this content for all.

Background of the ANDP

The ANDP is a collaborative program that is digitising historic Australian newspapers published between 1803 and 1954. The aim of the program is to develop one national access point for all Australian digitised newspaper content. Traditionally people wishing to research Australian newspapers have had to go to their library and browse through reels of microfilm, but today the Library is providing an online service that allows people anywhere, anytime to access these newspapers via the internet. The service is free to use and supports full text searching across every page of every newspaper in the service.

Early Australian newspapers

By way of background, early Australian newspapers are one of the most important resources that provide contemporary accounts of how the colonies were governed and of key historic events that shaped the nation. They reflect the day to day lives and circumstances of early Australians and are a significant record of the social, political, economic and cultural issues of the time. This is reflected not only in the written articles but also the images, advertisements and

even the headlines and layout of the newspaper. It is for these reasons that newspapers are heavily used to support historic enquiry.

Australia's first newspaper was the *Sydney Gazette and New South Wales Advertiser*, first published on Saturday 5 March 1803. It was a government gazette published by authority of the Governor of New South Wales with the important role of distributing official announcements, shipping news, excerpts from foreign newspapers and local social news. In each of the other Australian colonies, the first publication was also a government gazette. By the end of the 19th century a number of metropolitan, provincial and suburban newspapers were being published and weeklies were starting to appear. These newspapers all played an important role in reporting news from abroad as well as the recording of Australian daily life.

Access to historic Australian newspapers online

The Australian Newspapers Service¹ is the public search and delivery interface, developed to provide free online access to select out of copyright Australian newspapers. Initially one major daily newspaper from each Australian State and Territory (of which there are eight) was selected for digitisation; however as the Program has progressed the number of titles has been increased to over 90². The selected titles are being digitised from the date on which they were first published, through until the end of 1954, or when the newspaper ceased publication. From 1955 copyright will apply and digitisation of newspapers published after this period may be undertaken in the future if permission is obtained from the relevant newspapers publishers.

The Program aims to deliver over 4.4 million newspaper pages consisting of over 44 million articles by July 2011. The National Library of Australia is leading and managing the Program with the Australian State and Territory libraries contributing the source microfilm for digitisation. As at 10 February 2010, the Program is delivering 1,230,000 pages consisting of 14,067,000 articles, through the Australian Newspapers Service.

Through the Australian Newspapers Service the Library is providing access to every article, advertisement and illustration on every newspaper page being digitised. Users can browse the newspaper pages or search across the full text of the articles. Key users of the service to date have ranged from academics and family historians to social and economic researchers to school students.

As well as the selected newspaper titles being funded by the National Library of Australia, the Vincent Fairfax Family Foundation provided the Library with AUD\$1 million in additional funding to include the *Sydney Morning Herald* in the Program. As Australia's longest running daily newspaper, this nationally significant title is an important and very welcome addition to the Program.

ANDP digitisation workflows

In terms of the entire newspaper digitisation workflow, from creation of the digital image through to delivery, there are a series of steps which are undertaken.

Identify and locate selected newspaper microfilm.

In order to digitise and make a large volume of newspaper content available efficiently and cost-effectively, the Program is creating digital newspaper page images from microfilm versions of the selected newspaper titles. As the National Library does not own the majority of the microfilm required for the digitisation process the microfilm is being sourced and borrowed from the Australian State and Territory libraries, as well as a Sydney microfilming bureau, who owns the microfilm for a number of Australian newspapers.

Digitisation

The creation of the digital newspaper page images from microfilm is undertaken by a panel of external contractors. To date, over 3.2 million digital newspaper page images have been created.

Quality assure digital images

Once the microfilm has been scanned and the digital newspaper page images created, they are delivered to the Library where they are quality assured. This process involves checking that the images are cropped, oriented correctly to reading view and have been de-skewed in line with the Library's specifications³.

Content Analysis and Optical Character Recognition (OCR) processing

Once accepted, the digital newspaper page images are then sent to a panel of contractors off-shore, where content analysis and OCR processing is undertaken. This is the most complex part of the entire workflow and involves:

- Zoning each page into areas, that is, identifying each individual article and illustration on the page as well as other elements such as the masthead;
- Identifying and linking those articles together that continue across pages and linking any illustrations to the relevant article;
- Applying a category to each identified article. The ANDP is currently using the following article categories:
 - *News*, which is the default category;
 - *Advertising* - which includes both classified and display advertisements;
 - *Family Notices* – such as birth, death and marriage notices so as to support research on people, and in particular, family history; and
 - *Detailed Lists, Results and Guides* – such as horse racing results, television and radio guides, crossword puzzles and lottery results. This category allows users to eliminate results and receives the lowest relevance ranking when results sets are returned.
- This processing step also involves converting the newspaper page images into a full text, searchable file using OCR; and
- Re-keying of identified parts of the OCR text for each article. As the quality of the OCR, or “electronically translated text”, can vary greatly, the article title, subtitle, author and abstract (first four lines) of every

News article is corrected. This means that articles are more retrievable and the results of keyword searching are more accurate for users.

PUBLIC AVAILABILITY

All completed and quality assured digital newspaper pages images are then made available through the Australian Newspapers Service.

Innovation in the ANDP

The Library sees the Australian Newspapers Service as being truly innovative and unique in the way in which it delivers digitised newspaper content and engages with the online user community. The Program has embraced Web 2.0 technology in order to provide a cutting edge service that allows users to interact, contribute and add value to the newspaper content. There is currently no other equivalent full-text newspaper service in the world that allows users to tag, add comments and correct the electronic translated text.

The management system to support the newspaper digitisation workflows (Newspaper Content Management System), as well as the public search and delivery system (Australian Newspapers Service) were developed in-house by the Library's IT staff using open source software. At the time the Program commenced there were no suitable systems in the open marketplace that would allow the Library to fully meet the objectives of the ANDP, so a 'build rather than buy' approach was taken. The Library examined how other newspaper projects are being delivered and where relevant, has applied similar methods. The Library also took a close interest in the way in which other successful online services were implementing Web 2.0 technologies.

To this end the Library believes that development of the service has not resulted in a traditional library database, but rather is providing users with innovative ways of exploring and enriching full text resources. The interface includes relatively standard functionality such as relevance ranking and clustering of result sets, but perhaps the most innovative feature is that which allows users to correct the OCR output or the 'electronically translated text'.

The OCR challenge

While OCR processing works well for documents with a modern, consistent typeface and standard format, the nature of historic newspapers with varying fonts and print quality, as well as high article density with little white space between text, means that OCR accuracy is often low. As the human eye is much better at reading text correctly than the OCR software, the Australian Newspapers Service allows and encourages users to correct errors in the electronically translated text. The contributions that users make in correcting the text add value to the service and improve searching for subsequent users. It should be noted however, that history cannot be changed with one deft keystroke, as the original OCR text is still retained and remains searchable in the database, in addition to the digital newspaper page image remaining a true surrogate of the original.

Since release of the Australian Newspapers service in August 2008, (the Service was first released in beta), the Library has built up a very dedicated user community who have been very active in making corrections to the OCR text - almost 9 million lines of OCR text have been corrected to date. For the Library the take-up of users in contributing to the service has been overwhelming, with many users saying that text correction is proving to be an ‘addictive’ or compulsive activity.

This phenomenon in the online world is known as crowd sourcing, which refers to harnessing the collective power of a group of people online to help solve a problem, or in this case, help improve the quality of the content in the Service.

Introducing the ANDP’s top 5 text-correctors

The Library was interested to find out more about these text correctors and in particular, their motivation. The top five text correctors have held their positions in the Top Text-Correctors Hall of Fame since the Service commenced. Below is a brief profile of each of the top five text-correctors:

Julie Hemenstall
Located: Sutton Grange, a rural town near Bendigo, Victoria.
Interests: Family history, local history.
Age and status: 31-45 (stay at home mum)
Text-correction topics: Focuses on a single newspaper title ‘The Argus’, the region around Bendigo where she lives and her family names. Always adds tags to articles as well as correcting text.
Time spent correcting: 15-45 hrs per week.
Why she does it: I enjoy the correction - it’s a great way to learn more about past history and things of interest whilst doing a ‘service to the community’ by correcting text for the benefit of others.
Motivation: The knowledge that you are doing something that will benefit future people that wish to access articles on their family history.
Is it addictive? Yes. A must do mission.

John F. Hall
Located: Melbourne, Victoria
Interests: Family history research.
Age and status: 61-80 (retired)
Text-correction topics: Attempted to correct/edit every 'Hall' entry in the Family Notices section of the 'The Argus' (Vic.), and is presently half-way through the approx. 7,500 entries.
What got you interested: “When I stumbled across the 1929 Death Notice of my great-grandfather (John Sinclair Hall), & found confirmation of an unknown

<p>family. Further related finds convinced me of its worth as a research aid.</p> <p>Motivation: I enjoy the 'work', and hope that my contributions will allow future researchers to also have a 'Eureka' moment, as I had.</p>
<p>Fay Walker</p>
<p>Located: Brisbane, Queensland.</p> <p>Interests: Family history research, prisoners of war, war crimes.</p> <p>Age and status: 61-80 (retired)</p> <p>What got you interested: One day whilst doing some research for my family tree I came across the Australian Newspapers Service and I thought well maybe I can do this, but the only experience I had ever had at typing was sending the occasional email to family or friends There were no guidelines in place on how to do the text correcting so I was able to experiment and edited a few articles, and from then onwards I was hooked.</p> <p>Why keep doing it: Text correcting and tagging has become a huge part of my life now and my main hobby.</p> <p>Is it addictive? Very addictive. Many of the newspaper articles are so well written and make such interesting reading that I have to pursue them through to the end, to discover the final outcome of the story. Motivation: Digitisation of the old newspapers is a really worthwhile project, and I want to contribute my little bit towards this great project.</p>

<p>Ann Manley</p>
<p>Located: Sydney, New South Wales.</p> <p>Interests: Family history research, architecture.</p> <p>Age and status: 52 (semi-retired) "Too busy text-correcting to hold down a full time job".</p> <p>What got you interested: Finding interesting stories about ship wrecks from the 1800's and the unexpected places that family names appear.</p> <p>Motivation: New content being added to the Service daily and the ability to correct text to help other researchers.</p> <p>Is it addictive? Yes, a challenge and I enjoy it.</p>

<p>Lyn and Maurie Mulcahy</p>
<p>Located: Brisbane, Queensland.</p> <p>Interests: Family history, local history, shipping.</p> <p>Age and status: 55-62 (retired couple)</p> <p>Background: Working together on research. Also transcribing shipping lists</p>

for <http://mariners.records.nsw.gov.au>

Why they do it: We are sick of doing housework! We enjoy both the Australian Newspapers Service and transcribing the shipping lists, as they complement each other. The Australian Newspapers Service gets you going on a particular theme (e.g. shipwrecks, of which there are over 16,000 references) and it is difficult to know where to stop.

Why keep doing it: Because it's addictive. It helps us and other people.

Motivation: It is fascinating to read about Australian (and world) history first hand from the newspapers, and to see how much some things have changed, and also how some things have stayed the same. So far we have been lucky in organising our overseas holidays so that we can get our 5,000 lines done for the month to keep on the Hall of Fame list.

Why do they do it?

The results of this very small survey of just five of the over five thousand users who engage in text correcting highlights the sense of community service, volunteering spirit and the desire to help others that comes through as a motivating factor. Most consider the contributions they make to be for a "good cause" and can see benefits for the Australian research community in the work that they do. In addition the sense of making history, being a part of history and recording history is also important to most Australians. Perhaps these reasons, more than anything else, are why text correction has been so successful in the Australian Newspapers Service.

Overall the Library found that the three main drivers for correcting text are:

- Helping to provide an accurate record of Australian History;
- Wanting to accurately record family names; and
- A sense of community service or 'social capital' where actions help other users of the Service as well as themselves.

Also evident is the trust and loyalty the Library has built with this online community. The users identified as our top ten text correctors have remained consistent since the Service was released as a beta or prototype in August 2008. So these users, as with the Service, have grown and developed over time. In order to reward the amazing contributions made by the top text-correctors, in January 2010 the top five text correctors were invited to the Library to receive public recognition with an Australia Day Achievement Award. Each of the top text-correctors was awarded with a medal and spent the day viewing special and treasures items from the Library's collections.

One of the most important lessons the Library has learnt is that engaging with users and building virtual communities has been just as important as providing the data itself. The Library's users want to part of an extended community and by providing them with a mechanism to contribute and a high level of trust, they have responded with commitment and loyalty.

CONCLUSION

I would like to conclude by making a few comments about the importance of the newspaper digitisation work being undertaken by libraries around the world. Libraries are now very much content developers and enablers.

Firstly, libraries have an access imperative – we want information to be easily accessible to users and free of charge. We also want to provide access to the full content and for this to be in context; for example to allow users to navigate from a newspaper article, to the newspaper page and then to the entire issue.

Libraries also aim to achieve ongoing access. We are instrumental in the development and implementation of digital preservation strategies to ensure that content we produce and make available today can be access for as long as possible into the future. Unlike content providers with a commercial imperative, we don't cease to provide access once the content is no longer commercially viable.

In addition, libraries are expert in listing, describing, cataloguing and providing metadata for collection content. We share and use this data to enable an understanding of where and what our collections contain, which is a critical element in enabling successful access to content.

Libraries also have a long history of collaborating and working together. The Australian Newspapers Digitisation Program is but one example of how libraries have successfully worked together to achieve enhanced access to digital content.

Finally, the Australian Newspapers Service has clearly demonstrated that users want to engage and be involved with full text newspaper data in new and exciting ways. The use of Web 2.0 technologies can enable this. Users have demonstrated a willingness to work towards the 'common good', to volunteer their time, energy, skill, knowledge and ideas and to be involved long term in a program of national historic significance⁴.

NOTES AND REFERENCES

1. The Australian Newspapers Service available from <http://newspapers.nla.gov.au/ndp/del/home>
 2. The complete list of titles is available at http://www.nla.gov.au/ndp/selected_newspapers/documents/ANDP_Titleselectmainlist_Nov_2009_000.pdf
 3. http://www.nla.gov.au/ndp/project_details/documents/ANDP_Work_Specification_microfilm_scanning_v1.2.pdf
 4. Holley, Rose (2009). Many Hands Make Light Work Public Collaborative OCR Text Correction in Australian Historic Newspapers. ISBN: 9780642276940 http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf
-

NEWSPAPER DIGITISATION IN SOUTH AFRICA

Patricia Liebetrau

ABSTRACT

Libraries and archives are increasingly utilising and exploring ways of using evolving digital technologies to enable their collections to be discovered, accessed and utilised in a Web environment. Users will increasingly search across repositories for resources, regardless of their physical format.

Newspapers as one such format, provide a rich sense of detailed history by recording events on a day-by-day basis down the years. However, newspapers have presented special challenges for libraries, such as storage and preservation due to fragility and size. Traditionally therefore, newspapers, unless microfilmed, have not been afforded a high profile in library holdings. Digital technology and changes to library collection development policies driven by user demand should encourage libraries to reconsider their stance on incorporating newspapers into their digital collections.

Despite the richness of history held in newspapers, South Africa has been slow in digitisation efforts in this regard. DISA was the first initiative to start with conversion of newspapers from microfilm and digitisation from physical issues. South African newspapers in the vernacular present special challenges for Optical Character Recognition (OCR) and has been one of the reasons for lack of digitisation. There are several other reasons for the slow uptake.

This paper will look at the state of newspaper digitisation in South Africa and will attempt to offer some insights and lessons learned from ongoing initiatives.

INTRODUCTION

Digital technologies present opportunities for new ways of presenting the archive enabling new interpretations of cultural history in any given time. These opportunities, however, are reliant to a large extent on skills in information technology, internet bandwidth considerations, access to computers and servers, human capacity and funds. DISA: Digital Innovation South Africa (<http://www.disa.ukzn.ac.za>) is an initiative, based at the University of KwaZulu-Natal, South Africa, with a core function of using digital technologies to provide online resources in support of scholarly communication, research and teaching. The principles of open access, open source and open content underpin the development of capacity building and skills training. DISA is committed to finding ways to overcome barriers, in an African context, such as cost of proprietary software, advanced information technology skills, institutional capacity and bandwidth issues, in the utilisation of digital technologies. This requires exploration of alternative avenues to arrive at innovative, lightweight solutions that are practical, relatively simple to implement. These efforts are essential in bridging the widening digital gap between the wired northern and not-wired southern hemispheres, and within Africa itself.

BACKGROUND

Durban is a busy port city on the east coast of South Africa, serving as a gateway to the road and rail transport system to the interior of the country as well as frontline countries to the north and north west of its borders. In the early part of the twentieth century Durban's development as a colonial city, was clearly reflected in styles of architecture, in dress codes, in modes of transport and others. A hundred years later, and the city has developed into a thriving multicultural commercial and industrial African city, with a population of around 3 million inhabitants. The local museums and archives in Durban hold rich resources about our local history, including a variety of newspaper collections. These resources are heavily consulted and used by researchers on a daily basis, resulting in deterioration of their physical condition. Furthermore the climatic conditions of hot and humid weather in Durban, as indeed in other parts of Africa, are not favourable to preservation of archival resources without climate control. Long term preservation of our cultural heritage resources in this environment is challenging and threatening long term access by researchers and other users.

In South Africa a new history curriculum is being introduced throughout all levels of education. This early period of our local history as evidenced by physical resources such as photographs and other archival documents provides a rich and varied resource. They moreover provide a context against which to refigure the archive and provide new insights into our past history.

Digital technologies are able to play an important part in not only helping to preserve such heritage materials but at the same time enabling the presentation of these resources in new and exciting ways to support new teaching and learning methods, such as outcomes based education.

DISA, as a collaborative venture, has explored many ways in which to utilise digital technologies to revisit these archival resources, draw together resources and present them with a fresh perspective in a manner that requires relatively little capital, information technology knowledge expertise and skill. Open international standards and *de facto* web standards such as XML are important ways to facilitate global collaboration and exchange of data – all important considerations in building digital resources suitable for an African context.

DISA initiatives

DISA has been at the forefront in South Africa in implementing digitisation to provide online resources and in so doing has increased capacity in the area of digital technologies and contributes to re-skill and new skills development in South Africa and further north in parts of Africa. Over the past 10 years DISA has achieved a measure of success in multimedia digitisation. Text based journals were the first to be digitised and complete runs were physically sourced from around the country and sometimes further afield to create a “virtual” full text based resource of hard-to-find journals published during the period 1950 to 1994 (our first democratic elections). Some newspapers were included in these initiatives and full runs of “community” based South African newspapers were digitised as part of a pilot study.

Newspaper digitisation – from microfilm to digitised scans

The Campbell Collections at the University of KwaZulu-Natal house many newspaper collections; some bound and physically preserved others which have been microfilmed over the years. A microfilm reader in the Reading Room is heavily used by researchers painstakingly going through page by page looking for articles of interest – no quick search and no short cuts to finding relevant information. So, in order to convert the microfilm to digitised images a Minolta MS7000 microfilm scanner was purchased and training was received to assist staff in the conversion process.

Several sample rates were used to test the best result and it was soon obvious that the original microfilming was not of a very high standard and had been microfilmed from bound issues. See figure 1 below



Figure 1: Page scanned from microfilm

The results obtained from the microfilm scanner were disappointing and not nearly as good as the results obtained from scanning the original pages on a flat bed scanner at 300dpi, 8 bit greyscale from the original unbound issue. See figure 2 below

Indian Opinion

Founded by Mahatma Gandhi in 1903.

Former Editor: Manilal Gandhi—1918-1936.

No. 1—Vol. LIX.

Friday, 13th January, 1961

REGISTERED AT THE G.P.O. AS A NEWSPAPER.
Price: FOUR PENCE

Transfer Of Power To Africans Inevitable

INDIAN PAPER ADVISES RHODESIA : COME TO TERMS WITH AFRICANS

—Indian Opinion—India Service

NEW DELHI.—Commenting on the resignation of Sir Robert Tredgold, the Chief Justice of the Rhodesian Federation and the unrest and rioting in Southern Rhodesia, the *Indian Opinion*, of Madras, wrote recently: "In fact the time has come for the European settlers in Rhodesia to recognise the inevitability of a transfer of power to the representatives of the Africans... The Southern Rhodesian European extremists have a great sympathy for the practitioners of apartheid but plans for a link-up with South Africa may not be easy of achievement in the world today. The only wise course is for the settlers to come to terms with African nationalism and that without any further loss of time."

The following is the full text of a telegram sent to the Government of the Union of South Africa on the resignation of Sir Robert Tredgold, Chief Justice of the Rhodesian Federation, in an event of great importance and one calculated to cause concern both to Sir Roy Welensky, Federal Premier, and Sir Edgar Whitehead, the Premier of Southern Rhodesia. Recently there has been serious rioting in Southern Rhodesia as a result of deep resentment among the Africans. Sir Edgar has decided to deal with it

himself. The Moslems Commission has pointed out how the Southern Rhodesia Land Apportionment Act is regarded by the Africans as "a major discriminatory measure." This reserves for the sole use of Europeans some 45 million acres of land (naturally the best, as in the case of Kenya's White Highlands) and prevents the Africans from occupying plots in the main towns of Southern Rhodesia. Thus in Southern Rhodesia there exists among

Indians Causing "Headache"

NEW DELHI.—The problem of the people of Indian origin in Ceylon was a "headache" for the Ceylonese Government and they would like to settle it said Mr. Maniyappa Sumanayake, the Island's Minister in charge of Home Affairs and Industries, at a Press Conference. He added: "We mean to take it up with your Government at a future date. Our Prime Minister might come here for a discussion." No date for the visit had been considered yet.

Mr. Sumanayake said that illicit immigration into Ceylon continued and there was "a very well-organised traffic" which was

Ela Gandhi Married In India

THE marriage took place in India on Friday last (January 6) of Miss Ela Gandhi to Mr. M. Rangobin. Miss Gandhi is the daughter of Mrs. Suchila Gandhi, of the Phoenix Settlement and the late Mr. Manilal Gandhi. She is also a grand daughter of Mahatma Gandhi.

Mr. Rangobin is a member of the well known Rangobin family of Inanda, Natal. The young couple will return and settle in Durban in a few months' time. Their forwarding address in India is: Care of: Mr. Arun Gandhi, Times of India Office, Dadral.

Figure 2: Page scanned on flat bed scanner from original

The conclusion therefore is that scanning from microfilm is and can only be as good as the original microfilm. Unless international standards have been applied (and these have changed over the years) in microfilm capture, the condition may not be the best. However it is sometimes the only recourse that we may have in scanning old and fragile newspapers.

The success rate of Optical Character Recognition (OCR) software was low when run on the microfilmed image in Figure 1 due to the low rate of successful word recognition. By comparison Figure 2 yielded a far higher success rate due to better word recognition and less manual intervention. OmniPage software was at this stage being used to perform OCR. We have since changed to Abbyy Fine Reader which seems to have a higher recognition rate.

DISA has digitised the issues of Indian Opinion from 1950 to the end of its publication in the early 1960s. A service provider did the scanning and due to the size and condition of the physical originals, a camera was used.

The OCR was performed only on the English text as we have not yet the capabilities to OCR the text in Indian languages such as Urdu, Tamil and Hindi. Full text searching is therefore only available on the English sections.

The Durban University of Technology (DUT) has plans to digitise the earlier editions from microfilm and make them available on their website.

TRANSLATIONS

Many of our South African newspapers are printed in the vernacular and South Africa has 11 official languages. Figure 3 below is an image from a newspaper written in isiZulu.



Figure 3: isiZulu language newspaper

Translation into English provides a challenge. Human translation would be an onerous and extremely expensive and time consuming task and at present no means of automated translation has been developed. The OCR “training” to recognise African languages has also not yet been developed. So whilst it would be feasible to digitise and present images on the Web, the search functionality would be limited to metadata only and that indexing created by humans. Automated extraction of metadata, particularly important at article level, would be useful.

METADATA AND ARTICLE INDEXING

Newspaper metadata created manually by human indexers is extremely labour intensive, time consuming and consequently very costly. Automated extraction of metadata is really the only option for large scale newspaper digitisation efforts. Recognition of individual articles by intelligent software helps to improve the search functionality and improves browsing.

DISA has developed a South African relevant thesaurus and has used keywords from the thesaurus to index all digitised resources, including newspapers. This allows searching on keywords across all types of material and improves the researcher experience when looking for something specific. Assigning keywords by a human is sometimes criticised as being subjective. Digital information managers should be aware of the need for objectivity and maintain an unbiased approach when indexing. The manual assignment of keywords at article level is a time consuming approach and would not be employed for large scale digitisation.

Another approach is to use the thesaurus as a keyword list of places and people

and this helps to inform the correction of OCR and improves full text searching. This approach has not yet been used in South Africa.

SPECIAL CONSIDERATIONS

Double page spreads present a unique challenge when scanning newspapers. Scan each page separately and the gist of the article is lost; scan the two pages as one page and it becomes difficult to read on a full screen and the page numbering is affected. DISA chose to scan two pages as one page for better readability. However I am unconvinced that this is the correct approach. Articles that are printed on more than one page and are continued on later pages in the paper also present a special challenge to ensure that the full article is able to be returned as a full article and not as an incomplete article, having lost its ending. Manual intervention was taken to ensure that all articles were complete. This however is time consuming and therefore expensive in terms of human time.

In order to keep digitisation costs of newspapers as low as possible it is important to keep manual indexing to a minimum. It is however important to have a semantic context which may require some human indexing. A solution which enables the best of both human and automated indexing is desirable but is not yet readily available in South Africa.

TECHNICAL CONSIDERATIONS

DISA has used open source software as far as possible. All content available from our site is made available on an open access basis and no charge is levied for access and downloading. Creation of metadata in XML format, the *de facto* open standard for the Web, will ensure long term archiving and preservation in a non-proprietary standard. Text Encoding Initiative (TEI) was used to mark up the text and capture the publication details of both the physical and the electronic versions in one xml file. We have also used the Dublin Core (DC) element set to capture metadata and have found that using the DC element set was more suited to capacity building as it required less learning and teaching time. In this regard it is also less expensive to create the metadata – an important consideration in South Africa and indeed in Africa. However quality control is required to ensure a consistent standard of input, check spelling errors and ensure complete and valid records.

Newspapers require large storage and server space. It is therefore preferable to create pdf versions on-the-fly for presentation on the Web but this does require a reliable and consistent bandwidth. The high resolution files are archived and preserved and low resolution files are made available for access and delivery on the Web. As with all digitisation, the initial costs are high but the reproduction costs are relatively low.

NEWSPAPER DIGITISATION IN SOUTH AFRICA

Large scale digitisation of our rich source of newspapers is not very far advanced at this time in South Africa. DISA, up to the present, has attempted the largest

scale of newspaper digitisation. The slow pace may be in part due to special considerations such as hardware and software but it is also attributable to lack of funding, guidelines at a national level for digitisation and lack of institutional capacity. The South African government, through the Department of Arts and Culture, has up to now not supported newspaper digitisation by committing funds for such a purpose. The National Library has made some attempts to digitise their rich collections but require large amounts of funding in order to complete the process. Private archives and libraries, such as the Campbell Collections, lack the capacity, funding and skills to digitise. Microfilm is still used by researchers to perform their research, as frustrating as this may be it provides the only alternative at present to using the fragile physical issues, which most archives and libraries are reluctant to allow into the reading room for fear of further deterioration.

CONCLUSIONS

Newspaper digitisation presents a wonderful opportunity to make available online a rich source of social history of our country, a memory of our culture. The contribution of newspapers to teaching, learning and research is also important. However this remains a challenge due to several reasons discussed above. DISA is currently hoping to continue digitisation with a wider selection of newspapers. A digital-newspaper-in-a-box solution that overcomes some of these challenges would provide a wonderful opportunity to increase the pace of digitisation initiatives and ensure that our rich and varied heritage is preserved.

NEWSPAPER DIGITISATION: A SILVER BULLET

Deborah Novotny

THE IMPLICATIONS OF DIGITISATION FOR THE PRESERVATION OF EXISTING NEWSPAPER CONTENT

This paper will discuss the proposition of newspaper digitisation as a 'silver bullet', a single solution to the complex challenges of preserving and making available newspaper content. It will examine the implications of digitisation on the care and preservation of existing hard copy and microfilm content by looking at three areas:

- The British Library's newspaper strategy and plans for a new state of the art storage building at Boston Spa (the British Library's document supply centre in the North of England)
- The decision to replace microfilm with digital as the library's preferred choice of preservation surrogate
- The library's microfilm legacy – managing the library's large, and in many cases, vulnerable collection of newspaper microfilm.

NEWSPAPER STRATEGY

The British Library's newspaper collection is housed in its own dedicated library building in Colindale, North London, complete with reading rooms and microfilm stores. The collection comprises over 52,000 titles housed on nearly 50,000 km of shelving. This represents newspaper content from all over the world, from the 16th century to the present day. The Library's Newspaper Strategy is designed to maximise the preservation of this content – and to enable ready access to it for all of our users. It is innovative and forward thinking; a modern strategy for modern times.

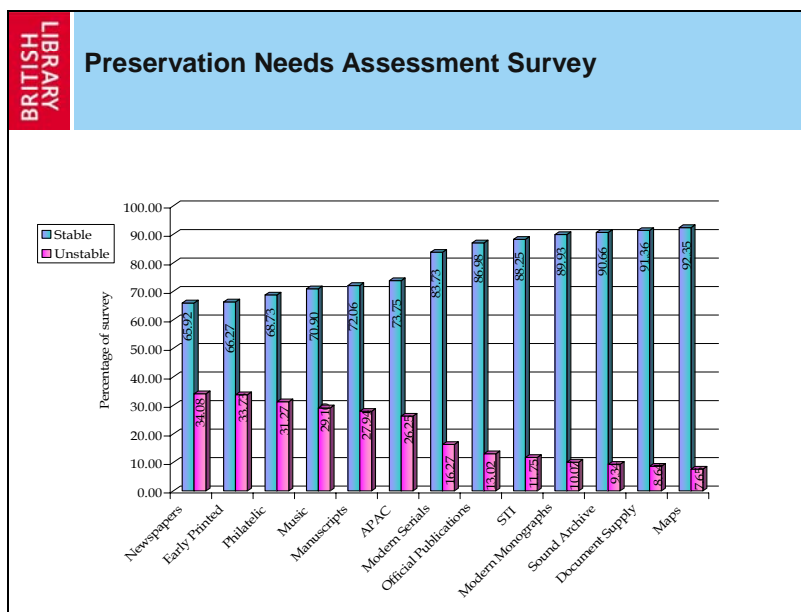
The systematic collecting of newspapers began in the 1820's when the British Library (as it is now) was still an integral part of the British Museum. Storage space was just as much an issue then as it is now, and by the end of the 19th century storage space was at a premium. One solution, proposed in the British Museum Bill of 1920, was to dispose of all of the regional newspapers pre-1837. But after major opposition to this, a building was purchased at Colindale in North London. This became the home of the English provincial newspapers and newspapers from Wales, Scotland and Ireland, and the first steps of a strategy to preserve and access newspapers were taken.

This solution was not without its drawbacks. London papers and overseas titles remained at the British Museum, creating a dual site. The building at Colindale had no facility for readers, and newspapers were transported from Colindale to the British Museum reading rooms in central London weekly, initially by horse and carriage, and then by van. And space remained a premium; this new building was itself full to capacity only twenty years later.

In 1928, the Royal Commission on National Museums and Galleries recommended that a purpose built newspaper library, complete with reading rooms, should be built at Colindale. This building - The British Museum Newspaper Library – was completed in 1932. It had a fully staffed bindery and a reading room for 56 readers and remains the heart of the current newspaper library currently in use.

Our Newspaper Strategy today is borne out of the same need to meet the challenges of storage and access as those faced by our predecessors. Continued growth and increased demand have meant that Colindale is now no longer a suitable option for the British Library in the 21st Century. Again, storage space is running out. The storage environment and infrastructure are not conducive to the long term preservation of the collections and the reading rooms, which remain as they were when they were first built, are frequently full.

In November 2009 the UK government confirmed a commitment of £33 million to fund the Library's Newspaper Strategy. This assurance enabled us to plan for the long-term preservation of the newspaper collection by constructing a dedicated newspaper storage building at Boston Spa. This will enable the closure of the library at Colindale completely mid 2013; the reading rooms are scheduled to close in August 2012, moving the entire collection to Boston Spa. Access will largely be via surrogate – both microfilm and digital – in a dedicated reading room at St Pancras (the target is 80% of reader requests being satisfied in this way). One of the Library's approaches to support its proposal to government for funding was to illustrate the condition and vulnerability of its newspapers. We were able to show from the results of Preservation Needs Assessment survey that the newspapers as a collection are the most at risk of all of the Library's collections, with over 34% being in unstable condition (15% of the collection so fragile that it cannot be used by researchers and a further 19% at high risk).



The new building will accommodate some 128 kilometre of newspapers in a low-oxygen, temperature 13C and RH 40%, high density, automated store, with a 3924 sq. metre footprint. The building will be similar to one completed on the site last year for the centralised storage of monographs (primarily) and low use material previously housed in multiple buildings across London. Retrieval for the newspaper storage facility is via an automated, computerised picking system, with the capacity to retrieve at 45 complete cycles per hour (both in and out).

The benefit of this new storage environment can be measured by the improved change in the newspapers' Preservation Index. (Preservation Index (PI) is a concept introduced by the Image Permanence Institute in 1995 to express the "preservation quality" of a storage environment for organic materials. PI has units of years. The higher the PI, the better conditions are for preservation of organic materials.)

The new building's environment will result in an increase in PI from 50 to 140 (years before deterioration is first noted). Or in other words the rate of deterioration would be half than what it was, providing a high tech, long term solution to the challenges of storing a large and vulnerable collection.

MICROFILM TO DIGITAL

An integral strand of the newspaper strategy, complementing the improved storage conditions, is the provision of accessing content through a surrogate. Traditionally, surrogacy has been driven by preservation with an access benefit. Consequently microfilm has universally been the preferred preservation medium. It is a cost effective, proven technology with reliable long term integrity.

However, the Library is an organic institution, constantly growing and adapting, seeking to exploit changing trends to maximise access to its collections and connect users with its content in ways which are both current and innovative.

Digitisation and the acquisition and presentation of content in digital formats sits at the centre of this philosophy - the balance is beginning to tip towards access driven surrogacy methods which can also satisfy preservation requirements. Microfilm is seen as old fashioned, colourless i.e. black and white and difficult to access compared to the glitz and ease of the digital. While microfilm enables the long term preservation of the original, and may itself be a source for creating digital content, it is not an easy way to engage users with content and its role is being questioned in the digital world. Consequently, Collection Care and the Preservation department is coming under increasing pressure to stop creating microfilm surrogates and create digital ones instead.

This is a complex issue in which some very real risks are bound up – both financial and reputational. There are serious questions to be asked- and answered - about access: not about the access benefits of digital *per se*, for these are undeniable; but rather how to make digital content quickly accessible to users in the existing frame work of the library. Digital brings with it inherent expectations (real and perceived) – that content can be accessed immediately, from anywhere, by anyone and is free. It is easy to create digital content, but if

it cannot be quickly processed and ingested; if its use is bound by copyright law or restricted by licensing agreements; if there is not the right infrastructure or enough infrastructure in the right places to make finding and using content quick and easy, then the access benefits are compromised and the preservation benefits are non-existent. So there are both risks and expectations to be managed in switching media.

There are questions, too, about storage. Digital storage itself is cheap and becoming cheaper, energy costs are not getting cheaper and what of the costs required to maintain the storage? And do these compare favourably with comparative costs for storing preservation copies on microfilm? In the current climate, long term value for money and the efficient use of funding cannot be allowed to be lost in the glare of digitisation's bright light. And there is the very real question of preservation and whether or not digital files prove to have the same long term preservation integrity as microfilm.

At the Library, the move to switch mediums is underway. From a Collection Care perspective, there are key criteria that must be met if this switch is going to be effective.

Some of the criteria, for example, we have considered when determining the change in a preservation copying policy are:

- The requirement for a trusted digital repository
- Proven metadata
- Accepted international standards for the production of digital surrogates
- Established workflow for records inclusion into a national/international digital registry
- Cost benefit of life cycle of digital asset management

In short, we find ourselves at a cross roads. We must move forward with a digital surrogacy programme, to support the Newspaper Strategy and its commitment to service 80% of requests by surrogate and to engage our users and promote the collections in ways that are relevant to them. Yet we have an obligation to preserve original content for future generations, and we can only do this if the surrogates we create are themselves able to be preserved and looked after for as long as possible. We must be confident that our digital assets can satisfy our preservation requirements, and that both access and preservation benefits are met equally in one format.

One solution we have considered that would enable us to push ahead with digital surrogacy without compromising our preservation requirements is to run both media in tandem. That is, create digital content for access and continue to use microfilm for preservation until the criteria for a full switch are satisfied. We can do this by using an archive writer – creating a digital copy first and then writing the digital file to microfilm (negative and/or positive).

MICROFILM COLLECTION & ACETATE

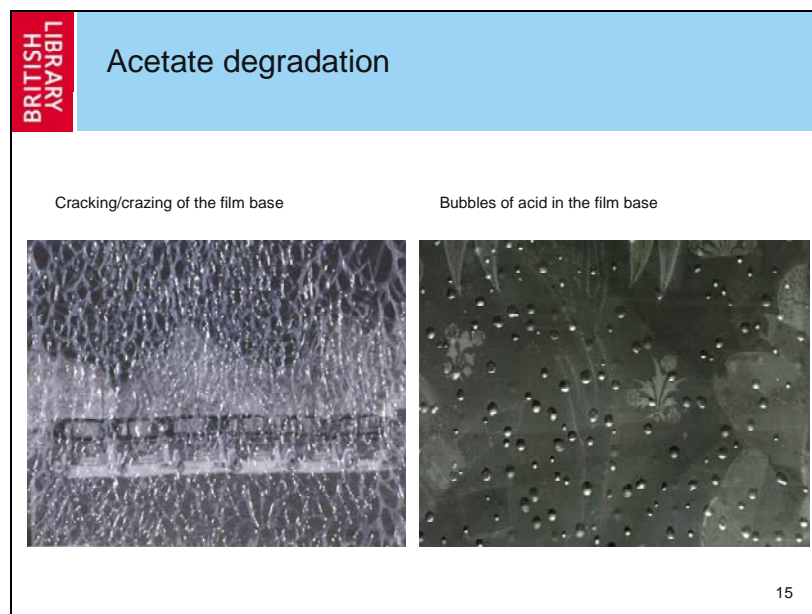
So, where does all of this leave microfilm? Does it still have a position of

relevance in today's access/preservation landscape?

The newspaper library first began on-site microfilming of its collections in 1950 and it has invested heavily over the years in the creation of microfilm – particularly of newspapers - and has built up a significant collection. On average we produce 4 million frames of new, preservation standard microfilm (one master negative, one duplicate negative and one positive) each year, at a cost per annum of approximately £900k. In total, the Library's collection of master microfilm of newspapers amounts to 28,202 x 1000ft cans and nearly 120,000 x 100ft reels – nearly 40million feet (rolled out and lined up end to end = 7,500 miles)

Microfilm - largely, and possibly only because of its physical format - may be becoming redundant as a first choice preservation medium in the face of digitisation (only time will tell if this is, in fact, the reality), but we cannot ignore our existing microfilm assets or their value in enabling the preservation of vulnerable hard copy, providing access to readers through positive copies and being an additional source for the creation of digital access copies. Consequently, we still need to provide resource to manage our microfilm collection, and make sure its long term preservation is secure.

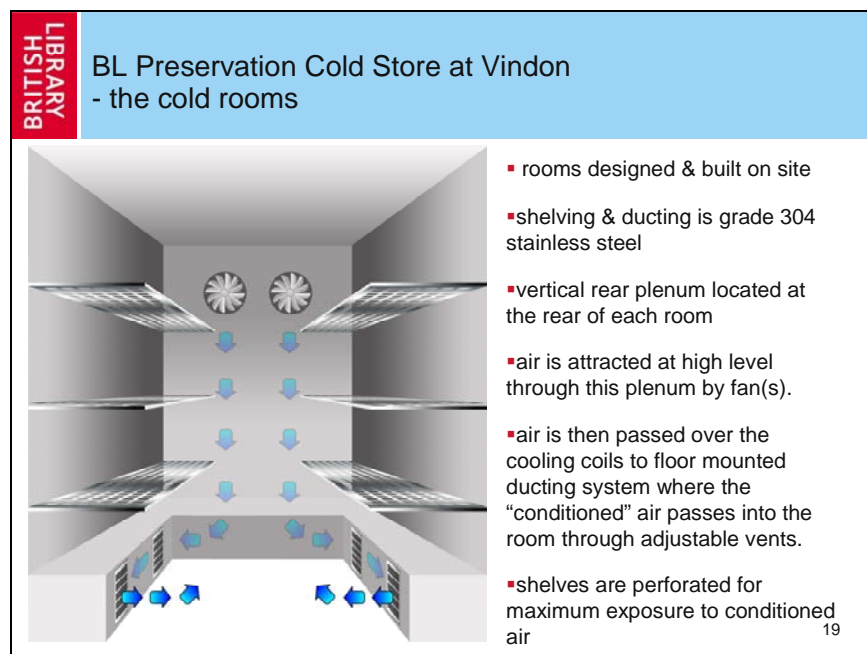
One of the major risks to our microfilm collection is posed by acetate film. This film base was introduced in the 1920s to replace highly unstable nitrate film and was most widely used between from 1930s right up until the early 1980s. It was marketed by Kodak as 'safety film', because it was much safer than nitrate. It was not, however, inherently stable. While it will not ignite and combust like nitrate, the film base will break down naturally over time, slowly releasing acetic acid, which then acts as a catalyst to speed up the chemical degradation. Acetic acid is commonly known as vinegar, which is why the advancement of degradation in acetate film is known as vinegar syndrome. The effect of this is reaction is brittle, cracking, degrading film.



Acetate film was replaced by polyester in the early 1980s— an inert and stable film base which is now the preservation standard. However, a lot of microfilming was created prior to this change, and British Library has estimated that – of its newspaper microfilm collection - about one third is acetate (13m feet; 2,462 miles).

In 2005, British Library took a proactive short-term strategy to this problem. We identified those acetate films for which we no longer held the original newspaper and, with £1m top-sliced off the Preservation budget over five years, we transferred them to a polyester film base.

As a long term strategy, however, this would not be cost-effective due to the sheer quantity of acetate in the collection and cold storage was developed. It is widely accepted that storing acetate film in cooler, drier temperatures is the best long term solution for acetate collections, as this inhibits significantly the rate of decay. In February 2009 after three years of research and consultation, the British Library awarded an external contract for the off-site provision of microfilm storage at 5 degrees C and 35% RH.



It is interesting that the company that won the contract is not from the heritage sector and has no history of heritage storage.

Vindon Scientific Ltd is a storage company, specialising in providing secure cold and cryogenic storage for the science and pharmaceutical industries.

A suite of 10 rooms has been built, each with its own independent monitoring system. The shelving in each room is perforated to allow maximum exposure to the condition air and also includes an acclimatisation room, so that film coming

in or out of the cold rooms is brought to the relevant temperature slowly and safely.

For the first time, the Library's master microfilm of the newspaper collection is united on one site. Moving the film to a much improved environment has resulted in the following benefits:

Acetate Specific

For our degrading acetate film (about one third of the collection), we have increased the time in which free acidity will double from about 10 years to approximately 200 years. For acetate not yet degrading, we have increased the years it will take for acidity to develop from about 80 years to about 350 years.

For our non Acetate specific, including polyester

We have increased the Preservation Index from 63 years to 488 and reduced the Natural Age Rate from moderate to very slow. (Preservation Index (PI) is a concept introduced by the Image Permanence Institute in 1995 to express the "preservation quality" of a storage environment for organic materials. PI has units of years. The higher the PI, the better conditions are for preservation of organic materials.)

And there have been other, collections management benefits of this move.

With every can in the collection having to be handled - some for the first time in many years - we have been able to do some much needed housekeeping.

- We have replaced over 6,000 damaged and/or rusted cans with new ones
- Every single item has been bar-coded for the first time and entered into a bespoke database.
- Online ordering system through the company's website for recalling cans when necessary and secure remote access to the database of holdings. This will be of benefit to the newspaper strategy in helping to determine what we have on microfilm and on what generation of film.

In an uncertain climate for preservation surrogacy in the near future, we have consolidated our master microfilm collection and are safe in the knowledge that it is recorded in detail and stored in an environment that will ensure its long term preservation for many more years to come.

But where does all of this leave the newspaper itself? Is it, like microfilm, a victim of its own format? Will the newspaper become an irrelevance compared to the value of the content it carries? And, conversely, will this lead to the hard copy originals actually being valued more as objects?

These are interesting and exciting times for the Library's newspaper collections - high density, automated storage; remote reading room access in the south via surrogate; digital preservation surrogates instead of microfilm.

Connecting all of these is digitisation.

But is it really the silver bullet, the solution to all problems? No, for all of the challenges of accessing and preserving newspaper collections have their own issues and risks and as some problems are solved others are created.

Digitisation is rather a tool – a powerful and exciting one – but a tool nonetheless to help us get the balance right between access to our newspaper collection and its long term preservation. It needs to be used judiciously in the right combinations with other tools and at the right time. Because, for our vulnerable newspaper collections and the information they contain, getting the balance wrong is simply not an option.

PRESERVATION AND ARCHIVING SOLUTIONS

Keith RajECKi

INTRODUCTION

Digital Repositories and Preservation Archiving have become increasingly important as institutions move away from the data management practices of deleting files and disposing of older books, magazines, and newspapers to a data retention mind set of preserving materials. Repositories help museums, national libraries, and educational institutions manage and capture institutional assets. This transformation in thinking has created a new set of challenges for these institutions financially, organizationally, and operationally. Past practices of deleting older files was an easy way to avoid the additional costs associated of increasing storage capacity. Removing older books and other materials from circulation in libraries made those shelves available for new materials.

Museums and national libraries are relying on digital repositories in order to preserve history .The ingest or digitization of historic photographs, films, maps, painting, and other items prone to deterioration allows for their long term preservation. The digitization and archival of these important historic artifacts is often times made available through the internet, enabling scholarly research, exhibitions, and other research and educational activities. Museums and national libraries have begun to leverage the digitization of these objects and technology to provide never before seen detail of manuscripts, paintings, and other artwork. Museums and libraries have even begun to develop three dimensional renderings of objects such as historic places, buildings, and people.

Higher educational institutions have turned to digital repositories for both academic and administrative purposes. A digital repository can hold a wide range of materials for a variety of purposes and users. It can support learning, research and administrative processes. Digital repositories provide a means of centralizing institutional data while providing greater levels of access and controls. Administrative digital repositories are primarily used for managing digital objects related to the administration of the institution such as student records, electronic invoices, and institutional policies. Academic repositories such as those in digital libraries manage books, papers, theses, and other works which can be digitized or were 'born digital'. 'Born Digital' objects are those that remain digital and are stored within the repository as they are created. Many organizations have begun digitizing their collections in an effort to provide open access.

The increase in the number of students taking online courses through the world wide web has also had an impact on the traditional service delivery of university libraries. These remote students are deprived of the vast majority of the resources traditional library have to offer. This has forced many university libraries to make their reference materials available online. This digitization effort has required a new business model for the university library. Libraries are no longer solely relying on the IT organization for support of their Integrated Library Systems (ILS). Librarians, IT Architects, and Systems Administrators

now work collaboratively as part of communities to develop and implement institutional digital repositories and preservation archives.

Digital repositories are also a key component to preservation archiving. The digital repository provides a mechanism by which the digital objects can be stored and thereby preserved. While the objective of the digital repository is typically related to collecting and disseminating digital objects, the objective of the preservation archive is focused on long term preservation of the digital objects. These preservation activities have expanded in scope to include audiovisual objects, datasets, presentations, text-based materials and research works. The preservation of these objects contributes to the re-use of digital content.

Purpose

This preservation archiving solution is intended to address the challenges of designing a digital repository and preservation archiving system regardless of types of digital objects needing to be stored. While the digital object types and properties should be considered when determining the storage components of the architecture, the object characteristics are not as important as the access and preservation requirements.

Digital Repository and Content Management Solutions

For the purposes of this solution architecture, digital repository refers to the system by which objects are stored for preservation archiving. There are a number of viable repository solutions available that provide the capability to store, manage, re-use and curate digital materials. Repository solutions range from the traditional commercially available content management systems to open-source alternatives. Repository solutions support a multitude of functions and can be internally developed or extended. The repositories must be sustainable and supportable in order for the underlying storage system to operate. The following repository solutions were highlighted for their ability to integrate into tiered storage architecture and their support for interoperability.

Oracle Universal Content Management

Oracle Universal Content Management (UCM) is a comprehensive repository solution providing the most unified enterprise content management platform for a broad range of purposes including document management, Web content management, digital asset management, and records retention. The extensible functionality of the Oracle UCM allows organizations to build and complement business applications. The Oracle Universal Content Management solution enables organizations to build strategic enterprise content management infrastructure for content and applications that helps reduce costs, easily share content across the enterprise, minimize risk, automate and streamline manual processes, and consolidate multiple repositories onto a single platform for centralized management. Through user-friendly interfaces, role-based authentication and security models, Oracle Universal Content Management empowers collaboration and preservation in a highly secure environment.

Ex Libris Rosetta

Designed in collaboration with the National Library of New Zealand and reviewed by an international peer group of recognized leaders and innovators, Ex Libris Rosetta enables institutions to preserve and provide access to the collections in their care, now and in the future.

- **Scalable:** Built on a distributed architecture that can support multiple, flexible server configurations, the system can scale up to manage digital collections of any size. By separating the permanent and working repositories, the system offers the security and redundancy required to keep collections safe.
- **Expandable:** Ex Libris Rosetta supports the functional needs of institutions that receive vast numbers of digital items on an ongoing basis. Batch loads, staging areas, and multiple deposit hierarchies enable users to manage the ingestion of new material efficiently and effectively.
- **Flexible:** Easy setup options allow institutions to configure the system to support their unique requirements and adhere to governing bodies' guidelines.
- **Standards-based:** Based on the Open Archival Information System (OAIS) model and conforming to trusted digital repository (TDR) requirements, the system provides institutions with the infrastructure and technology needed to preserve and facilitate access to the digital collections under their guardianship.
- **Accessible:** Preserved digital entities are delivered via viewers that are compatible with the constantly evolving format types and are supplied as part of the system and third-party applications. With built-in integration capabilities, the system can accept delivery requests from discovery and delivery applications such as the Ex Libris Primo® solution while ensuring that item-level access rights are enforced.

Fedora

Fedora is developed by the Fedora Commons non-profit organization as a platform for providing sustainable technologies to create, manage, publish, share and preserve digital content as a basis for intellectual, organizational, scientific and cultural heritage. Fedora is open source software built around a robust integrated repository-centered platform that enables the storage, access and management of virtually any kind of digital content. Content in Fedora can easily be accessed from the Web or by almost any software applications using available extensible application programming interfaces (API's). The connections between content items can be captured and stored in Fedora as semantic relationships describing both the linkage and its meaning.

Fedora is the first open source repository designed to work as part of an extensible framework of service components. This allows you to seamlessly incorporate Fedora into your organization's existing infrastructure. This extensible framework also allows Fedora to support trusted, secure organizational repository needs while supporting rapidly changing Web

services applications. Fedora's standards-based framework can incorporate the latest technology while keeping the content safe and accessible. Using this framework, you can easily add innovative technologies as services or plug-ins without compromising the trusted core.

DSpace

DSpace is an open source digital repository system that allows researchers to capture, store, index, preserve and redistribute digital data in virtually any format. More than 300 institutions worldwide use DSpace as their digital repository. DSpace provides organizations with an easy to use end-to-end solution for managing and providing permanent access to their digital works. DSpace was originally developed as a joint effort between MIT Libraries and Hewlett-Packard (HP). It is freely available to all commercial and non-commercial organizations under the BSD open source license. DSpace is designed to work out of the box and yet it also provides the flexibility to be easily customized to meet an institution's unique needs. DSpace Manakin provides a modular user interface layer, enabling institutions to design a unique look-and-feel that can be different for each community, collection and item across the repository. Manakin also allows the user interface to extend outside of DSpace into an existing Web presence.

DSpace supports multiple types of storage devices through a lightweight storage API. The storage layer currently provides support for local file systems, Storage Resource Broker (SRB), Amazon S3, or Sun SAM/QFS. New storage devices or approaches can be quickly integrated using the existing storage API's.

EPrints

EPrints is an open source software package for building open access repositories that are compliant with the Open Archives Initiative Protocol for Metadata Harvesting. It shares many of the features commonly seen in Document Management systems, but is primarily used for institutional repositories and scientific journals. EPrints was developed at the University of Southampton School of Electronics and Computer Science and is released under a GPL license.

EPrints is a Web and command-line application based on the LAMP architecture but has been ported and optimized for Solaris. Version 3 of the software introduced a (Perl-based) plugin architecture for importing and exporting data, as well as converting objects (for search engine indexing) and user interface widgets.

VTLS Inc. Vital

VITAL is a commercial institutional repository solution from VTLS Inc. designed for universities, libraries, museums, archives and information centers. Built on Fedora™, this software is designed to simplify the development of digital object repositories and to provide seamless online search and retrieval of information for administrative staff, contributing faculty and end-users. VITAL provides all types of institutions a way to broaden access to valuable resources that were once only available at a single location and to a finite number of

patrons. By eliminating the traditional limitations information seekers encounter, this technology grants access to materials for all authorized end-users, from professional researchers to recreational learners. Vital is a perfect solution for organizations looking for a commercially supported alternative to open source applications.

Islandora (Drupal/Fedora)

Islandora is an open source module for the Drupal web content management system written by the University of Prince Edward Island (UPEI) to allow Drupal to act as a web-based front end to the Fedora digital repository and preservation platform. The module also enables viewing and management of the Fedora repository objects, including functions such as ingesting, purging, adding datastreams, searching, and browsing collections.

Requirements and Design Criteria

Whether you are building a digital repository and preservation archive for historical preservation, to store data for business compliance, or meet the evolving business needs in higher education, a tiered storage architecture provides you with the most reliable and cost effective solution. Access and performance requirements are also important factors to consider when architecting your solution. Regulations often require that information be located and retrieved very quickly. If architected incorrectly, data searching and retrieval can be time-consuming and costly. Traditional tape only archival methods simply can not meet the access requirements of many of today's repositories and long term archives. Likewise, storing all the data on disk requires greater administration and is more costly. The proposed preservation archiving solution provides a proven solution with a balance between disk and tape storage hardware to support long term archiving.

Compliance

Higher educational institutions are facing complex compliance regulations for student records, financial, human resources, and donor information. While there are a variety of archiving products available, many products address specific applications, such as e-mail, while others enable broad archival for unstructured data, such as documents and audio/video files. This solution provides a tiered storage architecture that can be leveraged across a multitude of software solutions with a centralized policy based active archive capability. This preservation archiving solution addresses compliance requirements for non-rewritable, non-erasable format, the ability to verify automatically the quality and accuracy of the recording process, serialize the original and duplicate units of storage media, store separately a duplicate copy, and providing audit trails

Manageability

As with any multi-component solution, management complexities become a concern. The Sun preservation archiving solution design allows you to improve manageability of the technical infrastructure. The storage and archival functionality lead to improved manageability of the overall operational capabilities for the entire digital preservation organization. The software and

hardware components take advantage of open standards allowing fewer operators to manage greater storage capacity.

The ability to automate data management policies based on file attributes, enable you to manage data according to the storage and access requirements of each user on the system and decide how data is grouped, copied, and accessed based on the needs of the application and the users. This hands off approach to managing digital assets allows you to better utilize your valuable IT resources. The strategic placement of digital assets according to policy allows you to improve storage utilization and maximize return on investments by storing data on the most appropriate media type for the life cycle of the data while simplifying system administration.

Scalability

The ability for the entire technology infrastructure to expand and contract as services and storage requirements increase and decrease is increasingly important. The ability to dynamically reassign compute resources and storage as an organization moves from ingest to access drastically improves utilization. The Sun preservation archiving solution provides a highly scalable solution that takes advantage of server and storage virtualization to allocate system resources as needed. The components of the preservation archiving solution also enable additional resources to be added while the systems are on-line, resulting in live upgrades and transparent migrations.

Security

Digital archiving security requirements reflect concerns for long term access and preservation. Digital repositories and preservation archiving share the same security requirements as most enterprise applications with the added complexity of distributed object level policy based access. When long-term preservation spans several decades, generations, or centuries, the security of digital objects becomes critical. Open, standards-based access control, single sign-on and federation services are required in order to enable long term preservation while helping to control costs and minimize the risks of security obsolescence. The security solution must provide integrated user provisioning and identity synchronization services for securely managing identity profiles and permissions throughout the entire identity lifecycle.

Interoperability

Interoperability is the ability of software and hardware components to be functionally and logically interchangeable by virtue of their having been implemented in accordance with open standards. Interoperability of the digital repository and preservation archiving solution for institutional repositories is a complex problem. Interoperability is typically addressed between a specific software or hardware component. In this preservation archiving solution, interoperability is achieved among multiple software and hardware components through the use of open standards. In this model, different services and components can communicate with each other through open interfaces, and clients can interact with them in an equivalent manner. When repositories and digital objects are created in this manner, the overall effect can be a federation

of repositories that aggregate content with very different attributes, but that can be treated in the same manner due to their shared interface definitions.

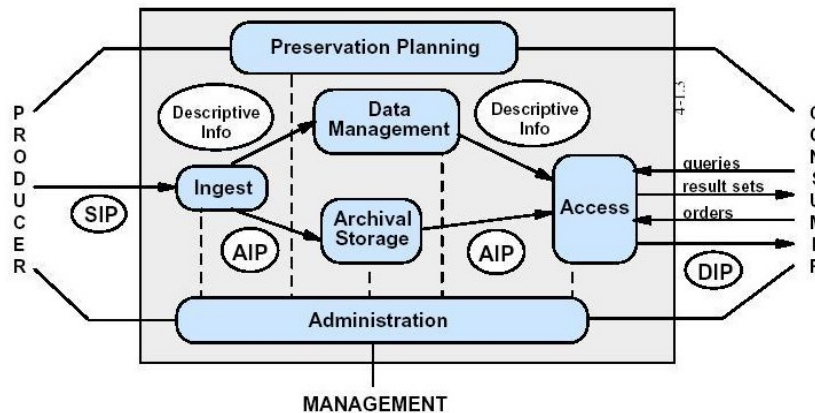
Logical Architecture

The primary functions of a digital repository and preservation archive is to support the acquisition, organization, preservation, and access to digital objects regardless of their format. The creation of functional requirements and identification of key policy issues for the digital repository are essential to building the appropriate architecture. The functional requirements for the Sun preservation archiving solution are based on the Open Archival Information System (OAIS) Reference Model and include ingestion of digitized and born-digital materials, metadata generation, data management, archival storage, access, preservation planning, and administration.

An OAIS-compliant repository is an organization of people and systems, which has accepted the responsibility to preserve information and make it available for a designated community. OAIS provides a conceptual framework to define the core requirements of the digital repository.

Open Archive Information System

While the Open Archive Information System (OAIS) model has become the de facto standard for preservation archives the design and implementation of a reliable long-term archive lacks adopted technology standards and design best practices. The preservation archiving solution provides a viable, cost effective, and reliable long-term archiving storage system based on the OAIS model and a combination of open source and commercially supported software and systems.



Services

The OAIS Reference Model has developed standard terms and definitions making it possible for those responsible for developing, implementing, and managing digital repositories and preservation archives to understand what is required to preserve and access information for the long term. The OAIS Reference Model illustrates the functions and information flows applicable to a digital repository archive. The major functions of the OAIS model are:

- Ingest - receipt and verification of records
- Archival Storage - secure storage of records
- Data Management - secure management of records
- Administration - management of internal and external relations
- Preservation - management of record integrity and security over time
- Access - provision of records in response to user queries

Archival Information System (OAIS)". CCSDS 650.0-R-1 – Blue Book.

Available at:

<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

The OAIS Reference Model does not prescribe any particular implementation, design, or implementation plan. Instead, it provides a framework for assessing the effectiveness and endurance of the preservation capacity of digital repositories. In OAIS an archive is "an organisation of people and systems that has accepted the responsibility to preserve information and make it available for a designated community".

The OAIS model describes systems with reference to information packages. An information package in the OAIS model contains content information, preservation description information and packaging information. There are three types of information package considered by the system:

- Submission Information Package (SIP), which conveys the information provided to the archive by the user and deposit system.
- Archival Information Package (AIP), which is the stored archival version of the information.
- Dissemination Information Package (DIP), which is the version of the information available to users.

Architectural Design

The Sun preservation archiving solution illustrates the integration of Sun software into the implementation of digital repositories and preservation archiving software. This preservation archiving solution delivers extreme levels of availability and offers proven enterprise-class scalability. The solution includes specific recommendations for hardware and software components that can help improve manageability, operational performance and efficient use of storage infrastructure.

Design Considerations

The first step to building long term archive storage architecture is the assessment of the business processes and defining the goals of your digital repository and preservation archive. Incorporating the business processes into your architectural design is crucial to the overall success of the long term archive. Documenting your organizations policies and procedures including data types, length of archive, access methods, maintenance activities, and technical specifications will increase the probability your archive architecture will meet

the business requirements.

A reliable long term archive is also dependant on the software components being open and support for interoperability. Storing, searching, and retrieving data is not sufficient criteria for a successful long term archive. A long term archive should incorporate open standards based software to ensure future support.

The overall storage system architecture addresses the physical storage components and processes for long-term preservation. Key components to address when architecting your long-term archive are security, storage, and application interoperability. The security layer focuses on the data access in order to ensure integrity and privacy. Storage addresses the placement of the objects within the various hardware components based on retention policies. Application interoperability is the systems and applications ability to be backward compatible as well as the ability to support expanded system functionality.

When designing your digital repository or preservation archive system it is important to understand the needs of the users of the system. Users are not limited to those who will be accessing the repository or archive looking for objects, but includes those who will be ingesting objects as well. Your users may consist of students, faculty, researcher, or even the general public. Each of which may have different access needs. These needs will influence the server requirements of your access tier as well as the performance requirements of your search and data retrieval. You must be able to define your acceptable levels of retrieval response times in order to ensure your objects are being stored on the most appropriate storage device. High speed disk systems will provide you with faster data access compared to tape library that may need to search and mount media prior to retrieval.

Funding is also an important consideration when planning your digital repository or preservation archive system. You must consider the operating and upgrade cycles of your infrastructure in addition to the initial acquisition costs. This will prevent you from implementing a solution that is either too costly to maintain or requires drastic re-architecture as a result of the growth of the repository. This solution takes advantage of low cost storage combined with open standards that lower your total cost of ownership.

As we discussed earlier, this solution supports a wide variety of content types. When planning your digital repository or preservation archive, you should consider the various content types you will be required to support. You may want to begin evaluating and planning different preservation policies for different content types. Not all content has the same preservation requirements or value. Flexibility of the tiered storage architecture allows you to expand and contract your individual storage tiers independently as your content storage requirements evolve. Here are a few examples of some of the content type you may be consider digitizing, ingesting, and preserving in your digital repository:

- Manuscripts
- Books
- Newspapers
- Music, Interviews, and Video

- Web Documents and Content
- Scientific and Research Data
- Government Documents
- Images
- e-Journals
- Maps

In addition to understanding your digital object types, you also want to consider the size of those objects as well as the total size of the repository. This will also allow you to forecast the growth rate of your digital repository in terms of the number of objects, object size, replication of objects, and total storage capacity. You will also want to establish and adhere to standard file formats when storing your digital objects such as tiff, jpg, or txt. It will be important that these file formats can be read by the applications that are available in the future when they are accessed from the repository or archive.

PRESERVATION ARCHIVING SOLUTION COMPONENTS

Open Storage

Open Storage refers to the systems built with an open architecture using industry-standard hardware. An open architecture allows the most flexible selection of the hardware and software components to best meet preservation requirements. A closed storage environment restricts the available technical components such as disk drives, controllers, and proprietary software resulting in higher costs and limitations to extending functionality. Long term preservation is directly dependant on the long term viability of the architecture and associated software components. Open standard solutions offer the most viable long term option with open access and community based adoption and support.

Hierarchical Storage System – Tiered Storage Solution

The Hierarchical Storage System, or HSM, is a key software element of the archive. The HSM provides one of the key components that contributes to reliability through data integrity checks and automated file migration. The HSM provides the ability to automate making multiples copies of files, auditing files for errors based on checksum, rejecting bad copies of files and making new copies based on the results of those audits. The HSM also provides the ability to read in an older file format and write-out a new file format thus migrating the format and application information required to ensure archival integrity of the stored content. The automation of these functions provides for improved performance and reduced operating costs.

The Sun StorageTek Storage Archive Manager (SAM) software provides the core functionality of the recommended archive solution architecture. SAM provides policy based data classification and placement across a multitude of tiered storage devices from high speed disk, low cost disk, or tape. SAM also simplifies data management by providing centralized metadata. SAM is a self-protecting file system with continuous file integrity checks.

Sun Storage Archive Manager addresses compliance by applying policies to files, copying and moving files based on those policies and maintaining audit information on files. SAM indexes files for searchability and writes multiple copies to specific media based on the compliance retention policies.

Designed to help address the most stringent requirements for electronic storage media retention and protection, Sun Storage Archive Manager Software provides compliance-enabling features for authenticity, integrity, ready access, and security.

Key Benefits of Storage Archive Manager Software

- Enforces retention policies at the storage level
- Software-controlled disks implement non-rewritable and non-erasable files
- Enables data integrity checking
- Provides flexible storage configurations

Storage Archive Manager software supports write-once read-many (WORM) files that are nonrewritable and nonerasable. Robust security features such as audit logs, user authentication, and access controls, combine to help safeguard the integrity of the digital information. In addition, the critical metadata attributes cannot be changed.

Infinite Archive System

The Infinite Archive System provides a pre-installed and configured hierarchical storage solution for digital repository and preservation archiving. The Infinite Archive solution scales easily providing petabyte scalability. The Infinite Archive System provides a three tier storage system consisting of the following components.

- Working Data Set, Online, on fast Fibre Channel (FC) Storage (Sun StorageTek Storage Array)
- First Level Archive, Midline, high capacity SATA storage (Sun StorageTek Storage Array)
- Second Level Archive, Nearline, high-performance tape storage (Sun StorageTek Modular Library System)
- Remote Archive provides a further level of archiving, with remote off-site storage of archived tapes

The Infinite Archive System takes advantage of Sun SAM/QFS software to manage the placement and retention of the data to ensure the most cost effective use of your storage resources.

Modular Tape Library System

The Modular Tape Library Systems provide scalable solutions up to 56 petabytes and 70,000 tape slots. This makes them the ideal platform for tape archives for off-line or dark archives. The Virtual Tape Libraries VTL enable tape consolidation with the low cost, cartridge removability, and long-term retention capabilities. This tiered storage solution is managed by policies on the

VTL, so the overall solution reduces your labor costs for virtual and physical tape management.

The Modular Library Systems provide greater levels of reliability ensuring access to your data. The robotic mechanism maintains reliability regardless of the number of expansion modules and helps to increase the stability and predictability of backups. Redundant, hot-swappable components, such as power supplies and fans, minimize disruption. An advanced digital vision system automatically calibrates the library to reduce wear and tear on the cartridge, drive, and robot. Dynamic worldwide naming and firmware code uploads eliminate single points of failure.

CONCLUSION

The Sun preservation archiving solution is an ideal framework for any institution looking to deploy a digital repository and preservation archive. The software and hardware components of this architecture support the long-term maintenance of digital objects by storing the digital object information along with the object itself. These components are organized into a system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility. The components can be upgraded and changed out independently with overlapping technology life spans to support the long-term requirements of the repository. The policy based distribution of digital objects throughout the repository further supports fiscal responsibility and sustainability. Incorporating industry accepted conventions and standards such as the OAIS model ensures the ongoing management, access, and security of deposited materials.

The Sun approach of working with leading digital repositories to integrate directly the digital repository software into storage software enables trusted digital repositories to manage the integrity and authenticity of records. This provides streamlined mechanisms to validate assertions about trustworthiness and provide the preservation processes that implement the required control and management capabilities. There are multiple technologies available today that can be used to build a digital repository capable of maintaining the authenticity and integrity of ingested objects.

The components outlined in the preservation archiving solution can be combined into a single system to validate, store, audit, secure, and preserve digital objects. The tiered storage architecture provides the most cost effective solution for object repositories and long-term archives while supporting scalability. The extent at which those storage tiers are deployed is dependant on the access patterns and archival policies. Although this architecture is not intended to cover all business requirements, it can be applied in a modular approach to address specific business requirements where one or more tiers may not be feasible due to business or technical requirements.

Key Benefits The preservation archiving solution identifies key system components and processes that are required to achieve high service levels and scalability. It provides the following major benefits to educational institutions:

- Higher service levels — the architecture is designed to optimize service levels with redundant components and automated failover using storage virtualization and cluster technologies.
- Reduced cost — Virtualization technologies enable consolidated solutions with higher resource utilization and tiered storage helps customers avoid over provisioning or under provisioning their systems. Best practices for management can also reduce the cost of maintaining the solution environment.
- Faster time to delivery — Accelerates deployment by providing proven and tested configurations with simplified installation to be up and running almost immediately.
- Reduced risk — validated hardware and software configurations greatly reduce the risk of unforeseen problems in a production implementation.

REFERENCES

Stanford Digital Repository (SDR)

The SDR is a set of services provided by Stanford University Libraries and Academic Information Resources/Digital Library Systems and Services (SLUAIR/DLSS) to support Stanford's institutional stewardship of information in digital form. The SDR services focus specifically on maintaining integrity, authenticity, and readability of digital information over time; they provide a trusted environment for long-term digital information storage and preservation activities. The SDR is one of several SULAIR initiatives focusing on Digital Preservation. In simpler terms, the SDR is a type of digital archive for trusted long-term digital information preservation.

Trust over time

The most important function of any long-term service ((partner)) is building and nurturing trust 'maintaining the trust of depositors ((information stewards?)) is the SDR's foremost priority. The SDR's approach emphasizes three key, symbiotic principles: Security, Transparency, and Proof.

Security

Security of access is maintained and enforced by the SDR. The owner or custodian of deposited data decides who, under what circumstances, is entitled to a copy of their data from the SDR. Security of digital information is also maintained by the SDR. Information is encoded and stored in durable logical packages designed to prevent undetectable degradation or corruption, and current best practices are employed to all but eliminate the risk of unrecoverable corruption or loss. The SDR's design and operations are optimized to support this security mission. For example, all information is stored redundantly across different types of storage at different geographic locations. Also, the network of systems trusted to store information is available only to SDR staff; it is not directly connected to the Internet.

Transparency

All of the SDR's methods, processes, technology, business mechanisms, and public statistics are available upon request to all depositors, members of the Stanford community, and members of the Digital Preservation community for review and auditing. The SDR reports regularly to these communities as well with current statistics on preservation activity, data recovery events, and business viability. Depositors, however, retain control over the dissemination of statistics tied specifically to their deposited data.

Proof

When a copy of deposited information is requested from and delivered by the SDR, the SDR provides proof that the copy contains exactly the same information as originally deposited. While the SDR's technology and business mechanisms maintain the security and integrity of deposited information, providing proof on delivery actually demonstrates successful preservation. Proof provides the ongoing assurance necessary for long-term trust.

An Archival Environment, not a Delivery Environment

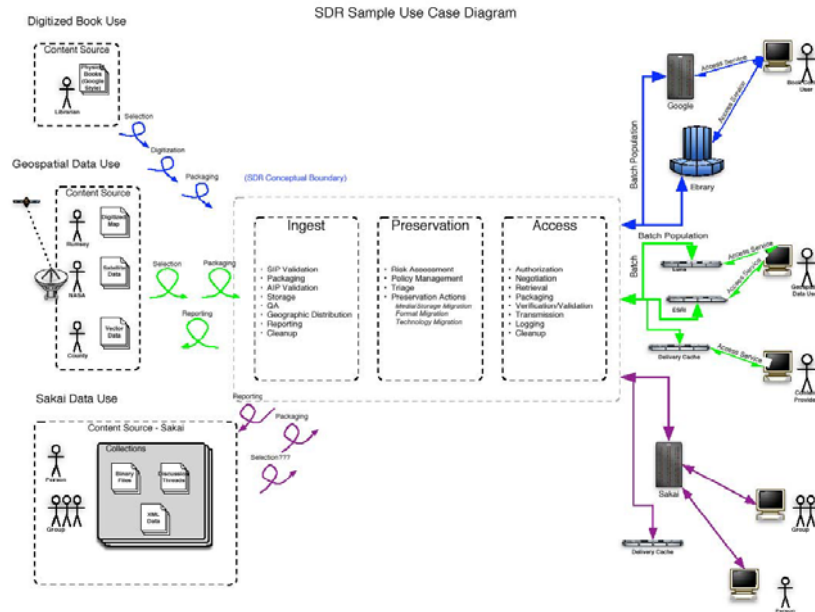
SDR services are designed for archival storage of information. They are tailored for periodic deposit and infrequent (i.e., neither daily nor weekly) delivery. The services do not constitute an efficient high-traffic delivery environment and should not be considered as a platform for daily end-user access. They are optimized for trust, not speed.

Currently available services

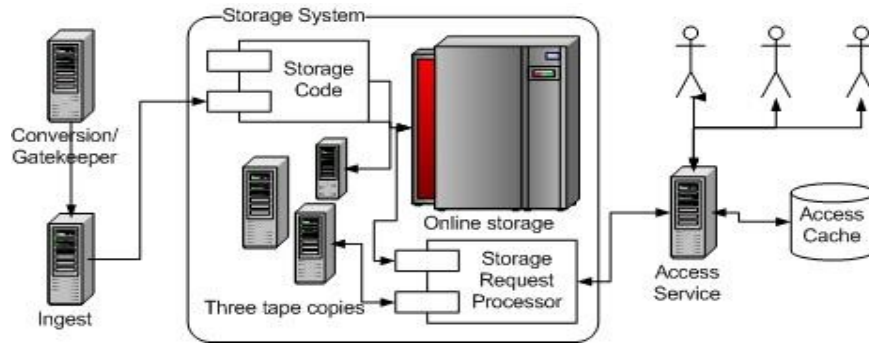
There is one service currently available: Bit-Preservation'. The SDR's bit-preservation service commits to stewarding ((preserving?)) digital files (aka 'bitstreams') and their metadata from any computer environment, ensuring the ability to deliver those files and metadata in current computing environments. For example, imagine storing files from an old Apple II computer, and being able to deliver those files in current environments such as Windows, Macintosh, Linux, and UNIX (and eventually other environments as they are developed and adopted).

Though bit-preservation for the long term can be quite valuable and is not a trivial activity, it does not ensure that files are easily usable in current environments. For instance, if a file from an old AppleII originally encoded an image, bit-preservation services can preserve and deliver that file to current environments but do not ensure that the current environment can render the file into an image.

The Stanford Digital Library is using Fedora and SAM/QFS in a tiered storage architecture environment for the digitization and preservation of institutional materials.



SDR High-Level Architecture



For more information

Web links for additional information

Web Site URL	Description
www.oracle.com	Oracle Universal Content Management
www.exlibris.co.il	Ex Libris Rosetta
www.vtls.com/	VTLS Vital
www.dspace.org	DSpace
www.eprints.org	EPrints
www.fedora-commons.org	Fedora

24 BY 7 DIGITAL ACCESS TO NEWSPAPERS: THE SINGAPORE NLB'S EXPERIENCE

Ngian Lek Choh

ABSTRACT

For several years, The National Library Board (NLB) of Singapore had been negotiating with the local newspaper publishers to grant permission for NLB to digitise the nation's newspapers and have them available online for remote access. The breakthrough came in July 2007 when the Singapore Press Holdings (SPH), the largest newspaper owner and publisher of our local newspapers in the four official languages signed a landmark agreement with the NLB to allow the NLB to digitize The Straits Times, its major newspaper in English and provide access from all of NLB's national and 22 public libraries from March 2009.

A second agreement was signed in July 2009 where the SPH gave NLB the approval to digitize all its other SPH newspapers and a number of its magazines published over the years. This agreement gave the NLB rights to provide not only onsite access to the full archives of the digitised newspapers, it also enabled the NLB to provide full text access via the web to the Straits Times newspapers from its first issue in July 1845 to 31 Dec 1989. The SPH provides born digital copies of its newspapers for issues from January 2007.

To improve search and find, the NLB obtained agreement from the SPH to let NLB put the first 50 words of every article in the newspapers and the titles of all articles out for internet search engines to crawl and index. This is so that users can find the articles from the comfort of their offices, schools or homes. Usage rose from around 30,000 page views in the first month of launch to about 90,000 today.

Challenges included the difficulty in bringing the separate columns of one article together when they are on separate pages. The team also found that there was no ready vendor to support NLB in the process of OCR for Chinese and Tamil newspapers.

BACKGROUND

Newspapers are a major source of information on a country's social history. This is no exception in Singapore. The National Library of Singapore (NLS) provides a newspaper service to its users since it opened its doors to the general public from the 1960s, and this was in the form of physical newspapers for current news and through microfilms for older issues of the newspapers.

For several years, the NLS had been negotiating for the rights from the media companies to allow the NLB to digitize the newspapers for easier access by the general public, from both within the libraries and outside. Under the NLB's Library 2010 plan, one of the major initiatives was to build a network of Asian knowledge assets to position the NLB as a library of global significance.

Included in this was a plan to develop Singaporean and Asian Resources to provide easier access to content about Singapore. Newspapers were identified as one of the key resources, and the library set a goal for itself to develop greater access to newspapers through digitization and push the boundary for remote access.

FIRST BREAKTHROUGH

The breakthrough came after more than two years of negotiations when the Singapore Press Holdings (SPH) gave the NLB permission to digitise its newspapers in July 2007. This landmark agreement provided NLB the rights to digitise its major newspaper in English called the Straits Times (ST). The SPH agreed to let NLB digitise all the past issues of the ST from 1845 to 2006. SPH also agreed to provide born digital copies of the ST from 2007. The agreement allows NLB to provide onsite access to the digitized newspapers starting from its first issue, from the National Library and 22 public libraries. The only exception to onsite access to the digital newspapers was for issues published during the most recent 7 days.

The NLB team initiated the project and completed the digitization of the first lot of newspapers from 1845 to 1982 during the first 18 months of the project. Tender was called to contract the digitization job to a company in India that could provide the service at a cost for the quality that was required.

SETTING UP THE PRODUCTION PROCESS

The first three months of the contract was a steep learning curve for both the NLB team and the vendor. The reason was that the NLB needed the vendor to ensure that the columns of each article are pieced together as a whole if they appear on different pages of the newspaper after the OCR process. The NLB also wanted to enable the users to search and find the articles by keyword and to be able to put the first 50 words and the title of every article out to Google to crawl and index, for greater access.

During the preparation period, the team found that the content of some articles were not pieced together correctly. The teams worked together with the vendor to improve the accuracy as this was an important outcome for end-users.

After a few months of preparations and testing the process for quality, the team was ready to go into full production from digitisation to OCR, to piecing together the correct columns of articles to enable search and find.

By March 2009, the issues of the digitized newspapers from 1845 to 1982 were ready for onsite access. Users were allowed to view and print the articles onsite. A fee was charged for printing. However, downloading was disabled.

As some of the content was put out to Google to crawl and index, users were able to find the articles when they search Google. When they find the article that they wish to view, they would be guided back through a link to the National Library website and advised to view the article onsite.

Though this was still inconvenient, it had removed the tedious process for users to find the right microfilm and to scroll the microfilm page by page to locate the exact article to use. This drastically cuts down the time to search and find the articles.

SOFT LAUNCH OF THE ONSITE SERVICE

In the first month when the digitized newspapers were put out, staff recorded over 8,000 digital visits in a month, and some 30,000 pages were read. This was without any publicity for the service. The team found that there was quite a bit of buzz in the internet space as users were telling each other about the service and how they had found it useful and easy to use.

FEEDBACK FROM USERS

Some of the comments received included the following :

- “This is great news for us! It certainly saves the tears and the backache from sitting through
- hours' of scrolling through microfilms in the icy-cold rooms at the library. Kudos to the library's
- digitalization efforts!” – Member of singaporeheritage Yahoo! Groups
- “With the new National Library Straits Times search engine, doing research like this has just been
- completely revolutionised!” – Member of singaporeheritage Yahoo! Groups
- “I am very excited that the Straits Times digitised newspapers are now online. I have found some
- matches for my ancestors and their vessels who sailed to Singapore in the mid-19th century...” –User from the United States.

Usage increased from about 3,300 microfilm users a day to 23,800 digital visits and 81,600 page views in Nov 09. This showed that there is a new group of users out there who needed information that they were not able to easily find before. Overseas users can now discover information about Singapore through this service.

THE SECOND AGREEMENT

Following the success in providing onsite access to the digitized ST, in July 2009, the SPH signed a second agreement with the NLB to allow it to digitize its other newspapers and magazines for onsite and remote access. This marks the beginning of a journey for NLB to digitise all the major national newspapers in Singapore in the four official languages (English, Chinese, Malay and Tamil).

The plan is to complete the digitisation of the newspapers in the next 4-5 years, and to allow anyone interested in Singapore newspapers to be able to search and

find articles published in any local newspapers using keywords, onsite and offsite, via the internet.

PARADIGM SHIFTS

The digitisation of newspapers provided a chance for the NLB to shift its perspective on what a newspaper service is.

Before this was possible, newspapers were provided as a service that was tied to a physical place. Anyone who wants to find a piece of information from any newspaper has to visit a library to locate the physical newspapers or microfilm and search each issue of the newspaper to find the article.

A story I told to my colleagues was the day we had a request from one of our senior government officials looking for an article published sometime in August 1960. Three reference staff took nearly a whole day to find the article as though we knew the gist of the article, we could not find the article in the August issues. It was found in another month and the staff had to look through all the issues almost for the whole year before we found it. This was not uncommon, especially when one is looking for an article that is not indexed.

With digitisation, the newspaper content becomes a mobile piece of content that can be moved from one place to another and with keyword search, finding relevant articles becomes so much easier. Although we have not been able to index the articles, the keyword search function has made searching for articles a breeze for most people as they reported the ease of finding what they need through the NewspaperSG service.

Once the article becomes detached from the physical device that carries it, the possible number of ways the library can expose and enable use of the article increases significantly. Where rights are given, the library can allow for anyone who wants to share the article by using the web links for them to freely share. Social media platforms can be used for users to share the content with their friends or communities with common interests. Even without the rights for access to the full article, the library is able to expose a small portion of the digitised content for internet search engines to crawl, index and enable search via the web. This increases the accessibility of the digitised content immediately.

It is amazing to see how digitisation completely changes the way the library can allow the user to search, find, share, use and re-use the content. The NLB will continue to find more ways to allow for more effective search, find and sharing of this content within the given rights.

LEARNING POINTS

Before the NLS started on this journey, we were very sure that we wanted to provide onsite and remote access to the newspapers. What we did not know was how to get to that point where there is a balance to satisfy the library's need to provide convenient access to newspaper content and the commercial interests of the media company.

As we developed the service, both parties learnt more about the possibilities of the service and better understand the objectives for the service. As the service is provided for users who use the content for reference and research purposes which is clearly stated in our terms of agreement for the service, and also, there are ways to protect the rights of the content such as the prevention of downloading and watermarking, SPH is more open to allow remote access to the content in subsequent negotiations.

We also added an additional step to inform the user that if they needed an article for reproduction for commercial use, they could go to the SPH's website and request for a copy for a fee. These steps gave the SPH the assurance that even if the user prints a copy of the article for use, they can only use it for personal reference and not to resell to a third party.

One other key learning point is that we should record every small and big agreement we made at every meeting and ensure that the notes of the meetings and what we had agreed to were agreed to after the meeting, before we took the follow-up steps. We learnt through the hard way that what we thought was understood was actually not understood in the same way that we had thought. We had to take a few steps back to clarify and confirm the understanding as we progressed. This took up valuable time.

CHALLENGES

There are a number of challenges that we face as we digitise and develop the services to enable access to the content. The first is to find suitable softwares that can complete the OCR process for Chinese and Tamil scripts. We learnt that there are softwares that can do OCR for Chinese script and we have tested two of the softwares. However, we have not put these into large-scale operation yet. It is too early to say if the digitised images after the OCR process would be of good enough quality for users to effectively find what they want.

For Tamil script, the team has not been able to find a software or vendor who is able to get the software for testing, and to see the results. This is a critical part of the process, as without the OCR part, users will not enjoy the necessary convenience of search and find.

The completeness, quality of the original newspapers and the OCR process impact the final digitised images and how accurate the search result can be. These are challenges that may not necessarily be resolvable by the team in the short term.

One other desired outcome is to find more ways for users to find and get the content even more conveniently, whether they are holding a laptop or mobile phone. We would like the content to be so accessible that they do not have to leave the digital space they are in, in order to find our content. This would make access more convenient and reduce the barrier to use.

CONCLUSION

We have been very fortunate to have such an enlightened partner in SPH, as they were prepared to let us provide remote access to newspapers that are more than 20 years old from anywhere outside the libraries. This was a great boon for users as they do not have to visit our libraries to find and use articles in the ST from Jul 1845 to Dec 1989.

Apart from reducing time in locating articles in microfilms, straining their backs and eyes, they can now also look for newspaper articles from the comfort of their homes and offices, saving time in traveling to the library too. Users from Italy, USA and elsewhere wrote to the library describing how convenient the service was, and how easy it was for them to find the information that they needed with the ease of keyword search.

For users who grew up in the Google and Yahoo age, the service was as easy as searching for information on these internet sites. Many users who were looking for information in ST and on Singapore did not know about the NewspaperSG service. They made their search on Google and sometimes find the articles in NewspaperSG. Once they find the article and they click on the link, they are brought back to the NLS website to retrieve the article and to use it.

We plan to develop the service further and bring the digitised newspapers right into the space of the user, wherever this might be, anywhere anytime, in the way they want it. This, we know will enhance the value of the library to the user.

REFERENCES

1. Ed King, British Library digitisation : access and copyright, IFLA Conference, Quebec, Canada, 2008.
2. Majlir Bremer-Laamanen, Connecting to the past – newspaper digitisation in the Nordic countries, IFLA Conference, Oslo, Norway, 2005.
3. National Library of Australia, The digital future of newspapers – report on the Australian newspaper plan survey, Nov 2007.

IMPROVING ACCESS TO DIGITIZED HISTORICAL NEWSPAPERS WITH TEXT MINING COORDINATED MODELS AND FORMATIVE USER INTERFACE DESIGN

Robert B. Allen

ABSTRACT

Most tools for accessing digitized historical newspapers emphasize relatively simple search; but, as increasing numbers of digitized historical newspapers and other historical resources become available, we can consider much richer modes of interaction with these collections. For instance, users might use exploratory search for looking at larger issues and events such as elections and campaigns or to get a sense of “the texture of the city... how the city was thinking.” To take full advantage of rich interface tools, the content of the newspapers needs to be described systematically and accurately. Moreover, collections of multiple newspapers need to be richly cross-indexed across titles and even with historical resources beyond the newspapers.

Keywords: History, Interviews, Modeling Events, Text Processing, User Interfaces

INTRODUCTION

Because an increasing number of digitized full-text historical newspapers is now available, we can begin to shift from searching them one title at a time to considering the way access to several titles can be coordinated. For instance, in the time frame 1880 to 1920 for Washington DC about ten different titles are available. Similarly, at the state level newspapers from several cities are being digitized and there should be synergies among them.

Collections such as the LC/NEH National Digital Newspaper Program (NDNP) collection are particularly useful for detailed text processing because they provide public domain OCR along with word coordinates and font identification in the META-ALTO format. However, because of factors such as the poor quality of the originals and the necessity of reproduction from microfilm copies, the OCR is of uneven quality. Moreover, there are additional aspects of the newspapers that are not coded in the ALTO files.

Allen et al. (2008) explored a “pipeline” processing model with a series of stages for creating article-level metadata. One step in this processing is segmentation of the page image. This segmentation is based on the identification of large fonts in the text which indicates a headline and the top of an article. The text of each of the segments was then categorized by genre and topic categories based on the standard developed by the International Press and Telecommunications Council (<http://www.iptc.org>). The metadata assignment was moderately successful for narrow categories which included highly distinctive terms but it was less successful for categories which depended on

nuanced language processing. Allen and Hall (submitted) explored regular news features which were not captured by the original set of genre categories.

In summary, it is easy to process some categories which are associated with distinctive keywords automatically but many other categories are less accurately processed even though most of the content is highly predictable based on factors such as its location in an issue. The more expert knowledge added the higher the accuracy. However, the amount of digitized newspapers is so large that it does not seem feasible for people, even groups of citizens engaged in collaborative correction, to make all the corrections needed. It seems unlikely that complete corrections can be accomplished by the automated process but automated processing can augment the capabilities of the human beings.

Text Mining and Modeling History

Text mining can be used to identify patterns in the text which should also be useful for improving the text processing. For instance, Allen et al. (2008) reported finding a seasonal pattern for the word “drought”. In fact, such regularities are easy to find and here we present rich data for two additional examples. These are drawn from the *Philadelphia Evening Ledger* from mid-September to December 31, 1914. As shown in Figure 1, occurrences of the term “Thanksgiving” peak at Thanksgiving and mentions of the term “Christmas” peak, unsurprisingly, at Christmas. Moreover, other terms related to the holidays such as “turkey” and “Santa” follow similar patterns.

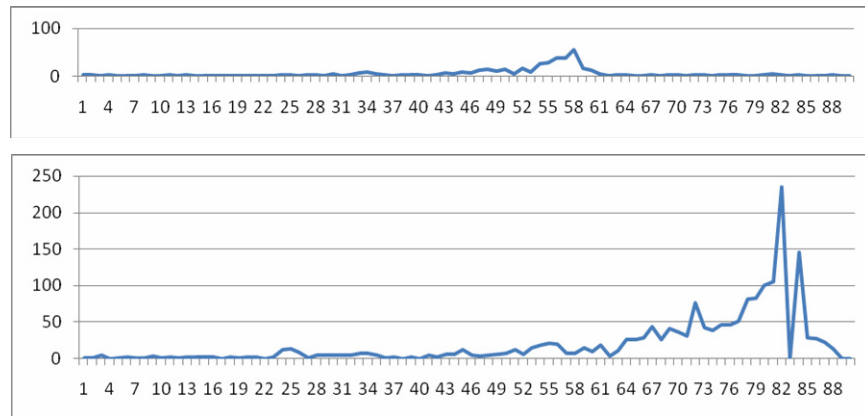


Figure 1: Frequencies for the term “Thanksgiving” (upper panel) and “Christmas” (lower panel).

Clearly, there are patterns in the data and they could be useful in making corrections in the text. For instance, they could be used to set weights for OCR corrections. However, to truly understand and exploit them it will be more helpful to develop models which account for and could even predict them. Thus, we want to move from simply observing such regularities to modeling them. Models could be based on many factors. For instance, we would predict very different types of news reports from a rural community than from an urban one. That is, we might develop what we could call “community models” (Allen et al, 2007). These would provide a unified framework for interrelating the people, places, organizations, and events that appear in the newspaper for a town or city.

Local reporting has a context within national and international reporting, even though readers at the time may not have direct access to non-local views. From a historical perspective there is, then, a requirement to enable the correlation of local news to the wider context and indeed to tracking the spread of news in times when local news readers did not have ready access to alternative news sources, such as telephones, television, radio or the Internet. A focus on local context also facilitates access to data that enables understanding of local social networks, social activities, sports, entertainment, community government functioning, advertising norms, and biographic data. These local contexts can then be compared across geography and time to analyze local and regional dispersion of news. In addition, to enhance understanding of the significance of search results, particularly from multiple newspapers, data such as local census data, economic data, or weather data, would be made available for overlaying on the search results. Moreover, relevant named-entities can be derived from many other sources. The “community models” that we propose incorporate these functions. Ultimately, these also need to be combined with newspaper models which encapsulate the editorial, stylistic, and production policies of each newspaper.

In this case, we are particularly interested in determining accessing civic processes, by which we mean government-related activities, which, of course comprise a substantial portion of the news. Figure 2 once again show data from the *Philadelphia Evening Ledger* for the fall of 1914. In this case, the data emphasizes terms related to the election of 1914 which was held on the first Tuesday of November. This is of particular interest because it shows a progression from the campaign to the election. Thus, we have identified a strong sequence of events by examination of word frequencies. This pattern is confirmed with terms such as “candidate” and “rally” for the days leading up to the election and the terms “election”, “votes”, and “voted” for the days surrounding the election. Obviously, these are very gross measures but they do clearly demonstrate the predictable evolution of events and, thus, we may think of them as being a model of events that generate news.

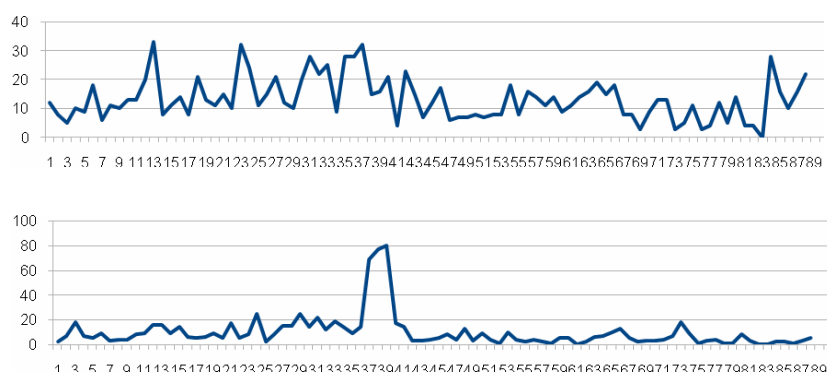


Figure 2: Frequency of the term “campaign” (upper panel) and of the word “vote” (lower panel). Note that term “campaign” is highest in the days leading up the election while term “vote” has a peak only for the few days surrounding the election.

Preliminary Results for Text Mining Concurrent Local Newspapers

While we have a fairly complete record for significant national events, we have a much less complete record for important events in individual communities. Such a record would be of interest in its own right in addition to being potentially useful for adding constraints for the text processing. However, identifying significant events from individual newspapers can be difficult because of the large amount of material to sift through. One strategy for focusing on significant events would be to compare coverage from different newspapers. There are a few cases where there are multiple newspapers digitized. In particular, the Library of Congress itself has processed a number of historical newspapers for the District of Columbia (DC). Of particular interest, there is a period of about three months in 1906 for which we have digitized samples of two major Washington DC newspapers: the *Washington Times* and the *Washington Herald*.

We collected the OCR output from the first pages of the newspapers during the period of the overlap. We cleaned that OCR by identifying only those words which were also found among the words in the Associated Press and *New York Times* portions of the Linguistic Data Consortium Gigaword corpus. That is, we tried to minimize the OCR errors by comparing the OCR text to words from a very large sample of English text. In addition, a stop list of the 500 most common words in English was also applied. Using the document-frequency measure from information retrieval, we identified terms which showed a striking change in frequency from their overall baseline. Finally, we found those words which showed that change of frequency in both Washington DC newspapers within a three-day window. These words are often indicators of distinctive news stories. Table 1 shows three examples selected from the output of this process.

Table 1: Examples of distinctive terms for news stories which appeared across two different newspapers in Washington DC on about the same day in late 1906. This technique allows us to identify news stories of particular significance because they appear in both newspapers.

Oct 29 1906	awful breaking bridge camden coach dempsey drawbridge heroism motorman picked submerged surface survivors thoroughfare trestle windows
Nov 18 1906	colon dillon hopes lacking princeton princeton's teams tigers yale
Dec 31 1906	ambulances awful belt coaches cotta crowded empty horribly identified mangled relief rescuers splintered takoma terra

These are preliminary results and much more work needs to be done to make them robust and useful. For instance, the mention of “colon” in the second line refers to the city of Colon, Panama and is unrelated to a main story detected about the defeat of Yale by Princeton in an American football game.

Toward a Historian’s Workbench

While dedicated researchers can and do exhaustively examine the rich resources of newspapers with microfilm, modern user interfaces for digitized materials should make the job easier for everyone and the barriers to entry much lower for beginning researchers. Thus, it is time to consider how historians might interact

with a much richer set of materials than they have previously been able to do. That is, we might think of developing a historian's workbench (cf., Toms & Flora, 2006). Robert Sieczkiewicz, the Drexel University Archivist, and I are conducting interviews with historians to determine the features historians would find particularly useful (Allen et al. 2010).

One set of issues concerns searching itself. For instance, one historian said. "The *[existing commercial online]* database is good for searching names, but broader topics are hard to research". Another researcher said she used newspapers to fill in gaps in research and corroborate information from other sources. Her exploratory searching included looking at larger issues and events such as elections and campaigns. That is, rather than searching on specific items, she used newspapers to find public opinion about issues such as changes in liquor license laws – to get a sense of "the texture of the city... how the city was thinking". That is difficult to do with simple keyword indexes of the materials.

Another set of issues deals with managing results from searches. One of the historians interviewed said "a log of all searches – this is a huge issue for me". When editing a book manuscript recently, she found it "hugely taxing" to find items she hadn't cited. Similarly, "searches lead to other searches", so she would like ways to see how searches are nested within each other and to return to earlier search results. She also asked for "a visual map telling you where you are in your search" as well a system that lets her easily use multiple windows.

Similar comments are also found in blogs on the Web. "Rachel", who describes herself as a doctoral student in history, presents a set of techniques for searching newspapers.

1. Have a List,
2. Find a thread of some kind,
3. Don't just use one newspaper,
4. Don't fall into the trap of only reading articles that your keywords throw up.
5. Use existing secondary literature,
6. Keep really, really scrupulous notes, and
7. Don't neglect the letters and the advertisements.

Taken together with the interviews, these may suggest that a user interface should have both flexible searching and also tools for annotation and management of the search results.

CONCLUSIONS

We have presented the value of extending text processing and text mining with modeling of underlying activities that are reported in a newspaper. Indeed, we can view a newspaper as a type of projection of, or perhaps a filter for, the community they are reporting about. In some cases these models can be based on predictable patterns. But, in other cases they are so difficult to predict that it is best to try to identify events as they arise. Thus, we also report results from

comparing the contents of two newspapers from the same city for the same time frame to find stories that are reported in both of them.

While we have explored many attributes of the historical newspapers, they have so much more rich material that we have barely begun to get a complete picture. Moreover, the application of findings such as these will be a huge undertaking. But, the result will be ready access to these exceptionally rich historical resources. Indeed, we envision digitized newspapers cross referenced with a wide range of other historical resources including primary sources as letters and secondary sources such as textbooks. Ideally, they would be woven into an automatically generated historical narrative.

BIBLIOGRAPHY AND REFERENCES

1. Allen, R.B., & Hall, C. (submitted) Automated Processing of Digitized Historical Newspapers beyond the Article Level: Finding Sections and Regular Features.
2. Allen, R.B., Japzon, A., Achananuparp, P., & Lee, K. (2007). A Framework for Text Processing and Supporting Access to Collections of Digitized Historical Newspapers. *Human Computer Interaction International*, Beijing.
3. Allen, R.B., Waldstein, I., & Zhu, W. (2008). Automated Processing of Digitized Historical Newspapers: Identification of Segments and Genres. In *International Conference on Asian Digital Libraries*, Hanoi, Vietnam, 380-387.
4. Allen, R.B., Zhu, W., & Sieczkiewicz, R. (2010). What to Do With a Million Pages of Digitized Historical Newspapers? Presented at the IConference, Urbana-Champaign IL.
5. Toms, E., & Flora, N. (2006). From Physical to Digital Humanities Library: Designing the Humanities Scholar's Workbench. *Mind Technologies, Humanities Computing, and the Canadian Academic Community*. Edited by R. Siemens and D. Moorman. Calgary: U. Calgary Press, 91-115.

NEWSPAPERS AS A DIGITAL RESOURCE NATIONAL INFRASTRUCTURE AND NEWSPAPER PUBLISHERS – IN FINLAND

Majlis Bremer-Laamanen

INTRODUCTION

The environment for news discovery in the Internet society, offers producers and custodians of newspapers, the national libraries opportunities and challenges.

1. In Finland the Ministry of Education is in charge of a project called the “National Digital Library” (2008-2011), with the aim to enhance access and long term preservation of electronic materials. Digitisation activities are promoted via project or other funding.
2. Newspaper presentation going digital at the National library
3. Structures for Licensing Agreements make solutions based on the law of Extended Collective Licensing possible. Access and license agreement structures for the library sectors exist e.g. via the Finnish Electronic Library.
4. First steps with newspaper publishers for digital newspaper retrieval.

The National Library of Finland has collected all print published in Finland since 1640. Accordingly the Library has been in charge of the national newspaper holdings since the first newspaper “Tidningar utgivne af et Sällskap i Åbo” was printed in 1771. The first Act on Legal Deposit from 1707 has been changed several times, it now also includes web-based newspapers.

Finland is one of the Nordic countries with a population of more than 5 million inhabitants. The main National Library is situated in the capital Helsinki and the Centre for Preservation and Digitisation in Mikkelä about 240 km from the capital. The Centre is one of the three departments of the Library. Its strength are in-house mass digitisation and preservation, conservation and microfilming activities. The Centre has been promoting development in these fields since 1990, when it was established.

THE NATIONAL DIGITAL LIBRARY (NDL) AND THE INTERNET ENVIRONMENT

The changing information society is the environment for news discovery, for newspapers from the past into future times.

The World Association of Newspapers and News Publishers WAN-IFRA promotes the freedom of press and journalism, research and business (www.IFRA.com). At the World Newspaper Congress in India, in Hyderabad in December, 2009, co-CEO of WAN-IFRA Timothy Balding summarised the digital advertising revenues to remain miniscule in comparison with print revenues. Search for new business models are the theme of today.

On a global scale according to the WAN-IFRA survey, newspaper circulation grew by 1.3 percent in 2008 and almost 5 percent during the last five years. The circulation grew in over 100 of 182 countries. The developing markets are rising and the developed world continuing downwards e.g. a 3 percent drop in Europe over the last five years. This is not the whole truth, because digital platforms have at the same time been embraced in the old world. Digital platforms, brand extension, social knowledge, attractive target groups and expansion to the young audience are much sought after. Here is to my mind, the possibility to find mutual interest to develop services together – between libraries, publishers and copyright holders.

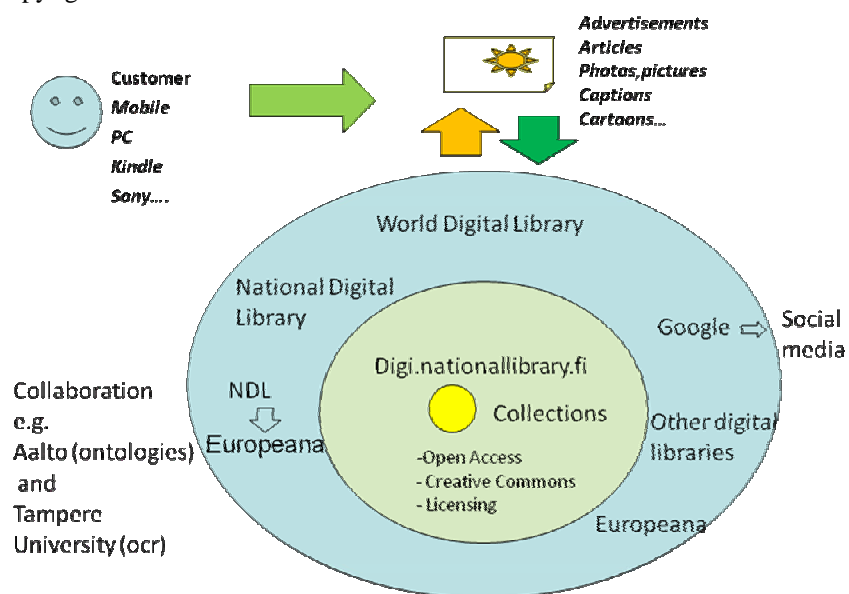


Figure 1: A draft of the digital library world today

The Ministry of Education in Finland in charge of the "National Digital Library" NDL project 2008-2011. The aim is to make the electronic materials of libraries, archives and museums accessible via information networks and to preserve them for future generations. The resources will be accessed via a common public interface in 2011. A long-term preservation solution plan will be ready in 2010, while the actual service will be available later.

The NDL is related to the European Digital Library and the 2010 concept of the European Commission: availability-access, long term-preservation and digitisation. It is of course related to the Finnish national development of electronic services and infrastructures, like "The National Information Society Strategy" (2007-2015).

NEWSPAPERS GOING DIGITAL AT THE NATIONAL LIBRARY

Access via digitisation

Newspapers are an important source of information in Finland. The biggest market in the world is India with 107 million daily sales. Again, in term of sales

per 1000 adult population Japan leads the world with 612, followed by Norway 576 and Finland 482. In terms of reach 91 percent of the Japanese read a newspaper daily when 76 percent do so in Finland.



Source: *Länsi-Savo*, 7th of February, 2010

Figure 2: First the cartoons! Newspaper-week at Finnish Schools in 2010

Microfilming of newspapers began in 1951 in Finland. Quality refilming for digitisation and optical character recognition have already been carried out for all newspapers dating back to the 1920's and is continuing. Microfilms are the platform for the digitisation of newspapers. The first digital Historical Newspaper Library was launched in 2001 as one of the partners of the Nordic TIDEN-project for newspaper digitisation. Today more than 1,5 million Finnish newspaper pages and 180 titles are digitised and searchable in many ways, e.g. freetext search, browsing, article search. You may like to have a glance at <http://digi.nationallibrary.fi>. All copyright free newspapers from the first title to the beginning of the 20th century are included in the service.

The use of the Historical Newspaper Library has grown by 42 percent in 2008 and 36 percent in 2009, now reaching 6, 5 million pages a year. This without any active marketing for the site – a thing we should do focusing on different user segments.

Today our challenge is to move from the copyright free papers to enabling access to the resources under copyright.

Funding and organisation

The National Library has continuously been funding part of our digitisation activities and external funding has made further progress possible. This concept began with the Nordic TIDEN-project beginning in 1998 and has continued

through the latest ones: Mass Digitisation of National Library Material 2007-2009 (Ministry of Education) with 1.8 million euros and special funding to boost employment in 2009.

These projects have enabled us to build customized tracking systems through the whole library. We are using Item Tracking and Scan Client software by Content Conversion Specialists GMBH CCS benefitting the processes for books, later for serials, newspapers and audio. Integrated management tools for statistics and project work flows are now in use. Our digitisation activities are now scalable for higher capacities according to available funding.

Within these projects we are continuously building competence for:

1. Library wide planning: Logistics for physical items
2. Implementation: Process for digital objects: network services and long term preservation of Metadata: developing, capturing through process customizing of the tracking systems (CCS, Item tracking, Scan Client)
3. Operational environment: scaling architecture and implementation
4. Digitisation of Newspapers and Journals and Audio.

The main process for digital objects consists of the following procedures: preparation, microfilming, scanning, post processing, access, archiving and back-up.

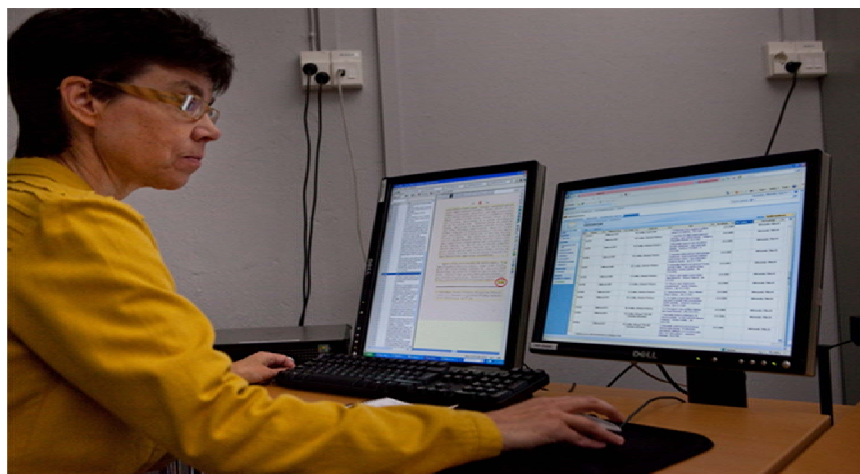


Figure 3: Riitta Alkkiomäki is making structural analyses of the content

For newspapers we have seen a change in the technical infrastructure that made it possible to proceed from the first phase of mass digitization to the second phase in 2007. The digitisation of newspapers has continued without interruptions. In the near future we are moving to the third phase for newspapers under copyright, where questions around accessibility will play an interesting part.

Metadata

The METS-format (Metadata Encoding and Transmission Standard) has been

used since 2004 for newspapers. METS is like a container. Within the METS, some of the following standards are used for newspapers and all of them are now used in the processes for books and journals.

- MODS and MARCXML for descriptive and bibliographical metadata (<http://www.loc.gov/standards/mods/>) (<http://www.loc.gov/standards/marcxml/>)
- MIX for technical metadata (<http://www.loc.gov/standards/mix/>)
- PREMIS for preservation metadata (<http://www.loc.gov/standards/premis/>)
- PREMIS for rights management metadata.

Newspapers in electronic format

The collecting of Finnish online content took off at the National Library of Finland in 2006, but it became part of legislation in 2008, when the new Act of Legal Deposit and changes in the Copyright law came into force. The National Library of Finland is obliged to deposit all the Finnish web content, which is available in the Finnish domains and servers.

Some of the Finnish newspapers are collected through the web page collections, twice a year, but the focus is not on the facsimile editions of the newspapers. One fourth of the newspapers published in Finland do also offer an electronic edition of the printed newspaper for their readers. These are available either for free or based on a subscription fee and they are available in varying formats, mainly in PDF.

As for the Centre for Preservation and Digitisation, one of the most prominent approaches in the future would be the retrieval of facsimile editions of Finnish newspapers in an electronic format and utilising this, printing the microfilm copy directly from the electronic edition of a newspaper. This approach is currently being investigated at the Centre.



Figure 4: The regional newspaper Länsi-Savo is interested to take part in developing Computer Output Microfilm at the Centre for Preservation and Digitisation. Ija Jaakonsaari is one of our specialists in the field.

GOALS AND STRUCTURES FOR LICENSING AND AGREEMENTS FOR NEWSPAPERS

The National structure for licensing is one important cornerstone for extended access to newspapers. Newspapers are a very much used legal deposit resource both as microfilms and especially when digitised. That is why the National Library is investigating possibilities to digitise and make Finnish newspapers available for research purposes until 1950 using the Extended Collective Licensing possibility.

Our goals are:

- to digitise all newspapers and about 8 million pages that have been published in Finland before the year 1950, by latest 2022
- to launch in 2012: the first newspaper and journals ensembles – the so called “baskets”
- to make agreements with newspaper publishers and the National Library to:
 - enhance digitisation of newspapers in copyright
 - enhance agreements for facsimile newspapers in electronic format
- to use the National Digital Library Public Interface as a platform for access to the holdings.

Note: newspapers under copyright will be available via the Legal Deposit Libraries in Finland – based on legislation 2008.

During 2009 working groups have been planning wider accessibility for newspapers both at the copyright society Kopiosto (including representatives of publishers) and at the National Library of Finland. Delegates from the Ministry of Education, Kopiosto and the National Library have a general interest to promote this collaboration. Some newspaper publishers have been directly in contact with our Centre promoting digitisation and access to newspapers.

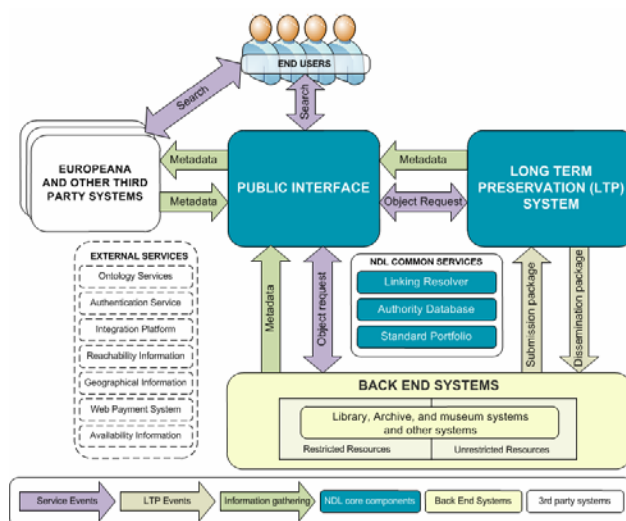


Figure 5: The National Digital Library Concept

Examples for extending access to legal deposit collections in copyright:

1. Solution based on the law of Extended Collective Licensing (16d§) The Copyright society (e.g. KOPIOSTO, TEOSTO) administrates licensing after getting permission for licensing from the Ministry of Education. The solution is based on legal requirements and statutory effects. This form of licensing is generally binding for copyright holders belonging to a specified group. The access to the material is specified separately.
2. Collective Licensing outside Extended Collective Licensing. Agreement between the copyright society and the organization-“Library”. The copyright society will respond to the possible claims. A funding mechanism is put in place.
3. Agreement directly with the individual holders of copyright. It is not suitable for Mass Digitisation.
4. This model is the guideline for the Finnish Electronic Library (FinElib)

ACCESS VIA THE NATIONAL DIGITAL LIBRARY

To whom would we primarily give access to the copyright based newspapers? The solution is of course dependent on the agreement of the publishers and the copyright society. A step by step solution could be useful, beginning with availability for research and study purposes.

Access and license agreement structures for the use by the library sector exist via the Finnish Electronic Library (FinElib). These structures could possibly be used or benchmarked when going further in the process.

The Finnish Electronic Library is part of our National Library's functions. FinElib is licensing electronic material for the library sectors (university, polytechnic, special and public libraries) in Finland. Using this structure, Finnish newspapers and journals could be licensed to those libraries and other organisations that are interested in the material.

The National Library will in the future be the custodian of the National Digital Library Public Interface in Finland. Newspaper holdings will be a most valuable resource here.

NEWSPAPERS AS A DIGITAL RESOURCE

Newspaper publishers are searching for new business models as Timothy Balding stated at the World Newspaper Congress in 2009. "As search engines take the largest slice of the internet advertising revenues, little is left for the content generators themselves". The world press advertising industry is 182 billion dollars while the digital advertising revenues of newspapers accounted for less than 6 billion dollars in 2008 and are estimated to rise to 8.4 billion in 2013.

New digital revenues are sought for and that is also what I think is happening in Finland. The published collections are the very essence of the newspapers.

These collections are also the focus of the National Library Legal Deposit collections. Both parties have the interest to give access to – and preserve the material. New forms of collaboration are sought for.

Here the structure and strategy of the National Digital Library can play an important part.

CONNECTING THE DOTS: HOW RESEARCHERS USE THEIR LIBRARY'S NEWS RESOURCES

Debora Cheney

Keywords: User Behaviours, Newspapers, Online Databases, Internet, News Industry

INTRODUCTION

Over the last decade libraries have been sorely challenged by the high cost of online databases and e-journals, a wide variety of information freely available on the Internet, and the changing expectations of students and faculty for a wider array of digital content. At the same time the newspaper industry has been challenged by changing economics, declining newspaper readership (particularly among young adults), increased reader demand for online news products, and a growing movement toward citizen "journalism." These changes also impact the use of library news sources. For this reason, libraries must learn more about how their news resources are being used and how trends in newspaper readership; the decline of the newspaper; changing research patterns and user expectations impact the use of news content provided by academic libraries.

This paper will identify trends and patterns in the use of library news products and sources by students and faculty at a large academic library. Database vendors data makes it possible to develop a picture of which news resources are used (and how much) and whether this differs or is similar to general news use and readership. Reported trends in student research behaviours also provide insight into the use of news content. These changes will likely impact future digitization projects, collection development decisions, reference and information services, and information literacy efforts related to news resources and content.

CHANGES IN LIBRARIES AND THE NEWS INDUSTRY

Libraries have made significant strides in understanding how researchers use library resources and what content and resources are used most frequently (Tenopir, 2003). Evidence suggests the availability of scholarly abstract databases, online catalogues (OPACs) and e-journals are significantly changing how students (Becker, 2009) and faculty search (Budd & Christensen, 2003) and their citation patterns. In response, libraries have begun to make collection decisions: changing what they purchase (for example, fewer print scholarly journals) and spending significantly larger percentages of their collections budgets on scholarly e-journals and online resources (Kyrillidou & Bland, 2009). Libraries have also sought to increase understanding, awareness and use of library resources by enhancing their website design (Fuller, et al., 2009), implementing link resolvers (Caudle & Schmitz, 2007), and focusing on the role

of the subject specialist in teaching information literacy skills (Barratt, et al., 2009; Head & Eisenberg, 2009).

In a similar manner, the news industry has been highly motivated to understand how users read and use their news products. A wide range of research and commentary on user behaviour related to newspaper and news readership (both online and print products) is available. Evidence suggests that readers (particularly young readers) access and “read” news differently (Vahlberg, et al., 2008); are often unfamiliar with traditional news forms; and expect news to be available 24-hours (Tapscott, 2008). Many “new media” (i.e., blogs, news websites, Twitter, and news aggregators) are impacting how news content is created, delivered, and managed.

On the other hand, libraries know relatively little about how and why students and faculty use the news content they provide. Some studies on citation patterns, (Kriebel & Lapham, 2008) provide some indirect information on the use of news resources. These articles provide some evidence news sources continue to be cited in student papers, particularly in the social sciences and humanities (Carlson, 2006). Some librarians believe access to more news content in *LexisNexis* does not necessarily increase the use of news resources (Leiding, 2005). Evidence suggests, composition and first-year courses still rely on news content but, within this context, faculty are struggling to teach students how to evaluate the credibility of web-based resources and may be relying more heavily on prescribing which resources students can use (Barratt et al., 2009; McClure & Clink, 2009).

USE OF NEWS RESOURCES IN AN ACADEMIC LIBRARY

Like many large academic libraries, The Pennsylvania State (Penn State) University Libraries has a long history of providing access to and preservation of newspapers. The University Libraries has participated in state wide efforts to preserve Pennsylvania newspapers and has traditionally provided a wide variety of news content in a variety of formats to support student and faculty research and teaching. At Penn State, *LexisNexis* has served as a flagship database ranking among the Libraries’ top five most frequently accessed full-text library database from 1999-2007. The University Libraries has licensed additional online news sources as they have become available and it is now possible to compare the number of searches and documents viewed across all online news sources and begin to identify changes taking place.

Although use of news resources may appear to be specific to Penn State’s faculty and students, it is possible to place them within the context of larger trends in the news industry and those challenges that are common to many libraries. Clearly the availability of a wide range of freely available news content is changing how many people access the news, particularly young adults who may never have read a newspaper. The Project for Excellence in Journalism has identified a power shift from ‘the individual journalist and away, by degrees, from journalistic institutions’ (Pew Research Center. Project for Excellence in Journalism, 2009b). This is taking place, as a greater number of news aggregators (i.e., *Yahoo News* and *Google News*) become available on the

Internet and news organizations are producing a larger share of their stories on new media platforms and some news is available *only* on the Internet.

Penn State's collection of online news resources can be compared individually, but it is more useful to compare them in groups (i.e., Aggregator Databases, Specialized Databases, and Historical Databases) as each group is used differently. In addition, evaluating the number of *documents viewed* (rather than *search sessions* or *searches*) provides a picture of how many articles were considered of possible interest to Penn State researchers.

Use of 'Old Brand' Newspapers

Penn State, like many libraries has come to rely on their Aggregator Databases to provide access to a wide range of news content and for wide geographic coverage. These databases have the largest potential user base and complement a library's existing (or non-existing) news collections. Because most Aggregator Databases have sophisticated search interfaces, they can also be used for a wide-range of student and faculty research. However, of the total news content and sources provided in these large databases, six newspapers overwhelmingly constitute the majority of use and the same newspapers are used consistently, regardless of Aggregator Database. In 2009 five major U.S. newspapers--*The New York Times*, *The Washington Post*, *The Chicago Tribune*, and *The Los Angeles Times*, and *USA Today*--accounted for 50% of all documents viewed [Table 1: 'Old Brand' Use Compared in Aggregator Databases, 2009]. Aggregator Databases account for 53% of all searches in all database categories [Table 2A: Database Use by Category, 2006-2009, by Searches] and 61% of all documents viewed [Table 2B: Database Use by Category, 2006-2009, by Documents Viewed]. Of these, 25% of the documents viewed are from *The New York Times* alone [Table 1]. Going further, 41% of documents viewed across all Historical Databases are from *The New York Times Digital Archive* [Table 2B].

This trend is consistent with larger trends for U.S. newspapers. News media insiders often distinguish between patterns of use between 'Old Brands' (or 'legacy') sources (i.e., *The New York Times*, etc.) and new media—(i.e., *MSNBC.com*). The 2009 annual report on *The State of the News Media* indicates 'Old Brands,' whether online or traditional formats, continue 'to have the greatest appeal' to readers (Pew Research Center. Project for Excellence in Journalism, 2009a).

Within the research and teaching context, these same factors likely influence the use of these newspapers; however, faculty influence and student research methodology are also important. Recently, Head and Eisenberg (2009) have provided insight into how students seek information for personal use versus course-related assignments. In the course-related research context, the study suggests students 'dial down their aperture' and rely on relatively few information sources, many of which they learned about from librarians in their freshman year or have proven effective in the past. Their approach, described by one study as 'power browsing,' (*Information behaviour of the researcher of the future: A CIBER briefing paper*, 2008) relies on *Google* and scholarly research databases (including *ProQuest*) as core finding tools—the focus is on 'efficiency and utility.' In addition, students, according to a variety of sources, are relying heavily on scholarly e-journals, based on their ready-access; their instructor's

expressed preference for these sources; and librarian recommendations. Both *Google* and scholarly journal databases allow students to ‘leverage’ the content they provide and help them to work most efficiently.

On the other hand, it is also possible that the use of ‘Old Brands’ in the research framework has much to do with their overwhelming visibility in multiple Aggregator Databases. Quite simply, as Penn State students focus on a limited number of methods and sources, ‘Old Brand’ titles above all others, are most likely to be presented in search results in the library database, *ProQuest*, used by the broadest swath of faculty and students. ‘Old Brands’ also appear in the other Aggregator Databases. Some research has shown ‘news portals’ (i.e., news aggregators) may also limit which news sources appear on aggregator websites most frequently (Bui, 2009)—thus increasing the visibility to a limited group of news providers.

Although a great deal of work has been done by Spink and Jansen (2004b) on the click-through rates for search engines, no research has focused on how effectively students and faculty are searching news databases. Xie and Wolfram (2009) have demonstrated the number of *search sessions* by ‘university/colleges’ in Wisconsin is declining and the ‘basic’ search feature is used most frequently in *ProQuest*. At Penn State the ratio of *searches* to *documents viewed* in Aggregator Databases is highest in the *ProQuest* database. Searchers appear to be either less efficient (7 searches to every article viewed) or are happening upon news articles as they cross-search multiple types of sources, including scholarly journals and magazines [Table 3 Ratio of Searches to Documents Viewed, 2006-2009]. *Newsbank* and *LexisNexis* searchers are possibly more purposeful searchers, with one search to each document viewed on average.

Barratt, et al. (2009) have argued ‘if students are going to use only one research tool, then these citations illustrate the importance of teaching them to search a multidisciplinary database that includes magazines, scholarly journals, and newspapers available in full-text.’ Certainly such databases contain the most ‘Old Brand’ titles. However, additional research is needed to understand how students use news resources for course-related research. Do students learn to effectively navigate and understand the variety of news sources available to them if they largely rely on such cross-disciplinary databases? Is it desirable that students limit themselves to one database or one search engine? Would a wider variety of news sources be used if students searched Aggregator Databases with deeper news content? Can librarian or faculty instruction have any impact on the search to documents viewed ratios? In addition, do students search differently if they are searching news contently only? More research is needed to clarify what the ratio of searches to articles viewed would be when the researcher is able to locate information effectively, rather than ‘Google’ or browse their way through information.

Use of State and Historical News Sources

While use of Aggregator databases and ‘Old Brand’ titles is nearly stagnant, both searches and documents viewed in Pennsylvania newspapers have increased as access to these titles has become easy in recent years. Penn State’s use of Pennsylvania news sources is consistent with use of Wisconsin

newspapers by Wisconsin libraries (Xie & Wolfram, 2009). The pattern of use merits additional attention. Pennsylvania news content use is spread across several different databases that provide quite different newspaper content in different forms, covers different years, and is likely meeting quite different research and teaching needs [See Table 1]. For example, in 2009: 16% of all documents viewed in Aggregator Databases were from Pennsylvania newspapers; 14% (26 titles) of those titles with 100 or more documents viewed were from Pennsylvania newspapers in *Newsbank*; and 35% (7 Titles) of historical Pennsylvania newspapers are included in the top 20 titles in *America's Historical Newspapers* (Readex); and, finally, while many fewer Pennsylvania newspapers are included in *PressDisplay*, one title (*The Pittsburgh Post-Gazette*) constitutes 5% of the total use of the top 20 newspapers viewed in *PressDisplay*.

Given the name recognition of many Pennsylvania titles, the quality of content, and years of coverage, this usage indicates state and local news resources continue to have value in the research framework that is quite different from newspaper readership patterns and news consumption, generally. For example, most news users acknowledge they would do no more than 'shrug' if their local newspaper were no longer available (Pew Research Center for the People and the Press, 2009). A recent study indicates a more complex use of local news sources and the role of local newspapers in that 'ecosystem' (Pew Research Center for the People and the Press, 2010); however, many (particularly young people) feel the quality of content and variety of other sources available for news (i.e., the Internet, television, etc.) makes local newspapers less useful for their own news gathering. Despite the larger trend, use of Pennsylvania news content in a variety of library databases continues to grow. Additional reasons for this pattern:

- At one time, researchers associated libraries with long term preservation and access to newspaper holdings and the continuing steady use of Pennsylvania newspapers may indicate some remnant of that association still exists.
- The documents viewed are unique state and stories that would not be accessible in 'Old Brand' newspapers. The 'half-life' of news stories on news websites, according to one study, is less than 36 hours (Dezsö, 2006). More study of the documents viewed in Aggregator Databases is needed to determine the 'half-life' of the news stories they contain— what years and stories or story types are viewed most frequently (the oldest content or the most recent)? How long does use persist?
- Aggregator Databases now contain as much as 10-15 years of newspaper content for many newspaper titles. For example, *The Pennsylvania Inquirer* begins coverage with 1981 and *The Centre Daily Times* begins in 1995. This constitutes a growing archive that is increasingly useful for a wider range of research. As many news/media companies begin to make their content less accessible (or charge for more content), the access provided by Aggregator Databases could play an increasingly important role in news-based research. Since the most heavily used Pennsylvania newspapers currently charge for access to

stories more than a few days old these newspaper's content are less accessible presently on the web, then say, *The New York Times* content, which at least *appears*, to be more accessible.

Meanwhile, a wide variety of scholarly research continues to rely on historical newspapers for research in many disciplines (Rettig, 2004). Today, historical newspapers are accessible in a variety of formats, but many newspapers may only still be accessible for some years on microfilm. Use of the University Libraries' newspaper collection on microfilm is sporadic at best. Requests of newspapers on microfilm via the University Libraries Inter-Library Loan (ILL) service continue at modest and largely static levels since 2002, while nationally ILL borrowing has increased significantly (Kyrillidou & Bland, 2009). It would be hard to predict which titles will be needed based on the titles requested; however, overwhelmingly, faculty and graduate students in the Colleges of Liberal Arts; Health and Human Development, Arts and Architecture, and Communications are initiating these requests. Requests are for pre-1924 years and requests for subsequent years decline as more content becomes available in online news databases or on the Internet [Table 4: Newspaper Requests from Inter-Library Loan (ILL), 2004-2008].

At Penn State, 22% of all documents viewed are from historical database compared to Aggregator Databases. [Table 3: Percent of Historical News Content Use Compared to Aggregator Databases, 2006-2009] Additional research is needed on how use of freely available newspapers at sites such as the Library of Congress Chronicling America differs from the use of titles contained in licensed databases. It is difficult to evaluate use of these resources until the usage data adheres to the Sushi guidelines. As more historical newspapers are searched in Google News, libraries should monitor closely the impact of that change on the use of licensed historical database content. Will this change increase or decrease use of a library's historical news sources?

Libraries still retain important newspaper holdings and have significant investment in news content. However, the archival digital divide and the role of Aggregator Databases is becoming a greater concern (Quint, 2009; Wiggins, 2002). As libraries move forward they must begin to consider how to re-insert themselves in the preservation and retention of today's news content, regardless of form and format (McCargar, 2005). If libraries do not preserve today's news content, it will greatly limit their ability to support the teaching, learning and research needs of their faculty and students tomorrow.

Use of International News Sources

Because students appear to limit the methods and sources they use for their research needs, this may partly explain why use of international news sources is also low despite the wide number and variety of such sources available in Aggregator Databases. In Newsbank, for example, 7% of documents viewed are from international sources. [Table 6: Use of International Sources for Newsbank, 2009 Documents Viewed]. There may be additional reasons faculty and students overwhelmingly view more documents from U.S. news sources than international news sources:

- Lack of language skills, although this does not explain the low use of English-language non-U.S. news sources;
- Database interfaces, for example, LexisNexis defaults to search ‘Major U.S. and World Publications,’ which includes no non-English news sources;
- Few international sources appear in the multi-disciplinary databases such as ProQuest; if *Le Monde* or *Le Figaro* were cross searched in these databases, would students view more documents from these titles?
- Such sources are used relatively little, overall, in teaching and learning; generally, researchers and students continue to use those news resources most familiar to them. If students are limiting the resources they use for their research, then international sources may be the easiest of all to avoid. Only instructors who specifically assign the use of such sources are likely to be rewarded (Darst, 1991).

Studies consistently demonstrate news readership must begin early and be encouraged as a habit (World Association of Newspapers). The same may be true of educating students and faculty about the role libraries serve in providing access to news content. Given the prevalence and widespread availability of news content on the Internet, news aggregators (i.e., Yahoo News and Google News) and ‘power browsing’, it is likely students are not relying on the University Library’s Aggregator Databases to ‘read’ the news. However, many newspapers are developing e-edition newspapers and PressDisplay provides access to thousands of these newspapers in many languages and from many countries. Students who use PressDisplay to ‘read’ rather than search the news may come to understand the role libraries serve in providing access to today’s news content.

At Penn State the top 20 e-edition newspapers viewed in PressDisplay, aside from U.S. sources, were European, Asian, and Business newspapers. The Sports category (i.e., sport section stories from a wide range of newspapers, as well as newspapers focused on sports) was also in the Top 20 ‘news sources’ selected in PressDisplay. [Table 7: News Use Online and In Room, 2009] This is a reflection of PressDisplay’s top page which features several categories, of which only the Sports category was used significantly. This pattern of newspaper readership is duplicated by readership of newspapers provided in the News and Microforms Library and at the University of Michigan (Freeland & Bailey, 2008). This may indicate there is still, albeit modest, demand for U.S., international, and business newspapers (either e-editions or newspapers in the library) and sports content if it is available on the day of publication.

Academic libraries have preserved and provided access to international news sources in an effort to support teaching and research, as well as to contribute to the diversity of their collections. However, use of this content overall is relatively small. Librarians may need to find ways to teach and market this content more effectively or work with database vendors to make core international titles more visible in their search interfaces and search results. In addition, libraries and database vendors may need to follow the lead of ‘news

recommendation engines' who 'have the potential to change patterns of news consumption by allowing readers to communicate both with each other and, indirectly, with news institutions themselves' (Thorson, 2008).

Use of New Media

Until relatively recently libraries have had little new media content, including television news, to offer their users. Today, *Newsbank* and *Vanderbilt Television News Archive* provide online access to current and historic television news programming, respectively. In addition, Aggregator Databases are including a wider variety of text-based media (i.e., blogs, newswires, transcripts, and news website content). News transcripts and wire services, while not new, are more readily available and as newspapers rely more heavily on wire service stories for national and international stories, access to this content separate from specific newspapers will be increasingly valuable (Weaver & Bimber, 2008).

At Penn State, the use of blogs, website content, and news programming video is still modest compared to 'Old Brand' newspaper content. Use is shifting from news wires to video and web content [Table 8: Use of New Media, 2004-2009]. Use of historical news programming video, as with all resources, increases when an instructor incorporated and required students to view historical clips in *Vanderbilt Television News Archive*, as in 2008 [Table 9: Use of New Media by Vendor, 2004-2009].

This change is consistent with reported trends. As the use of the Internet and broadband use has increased so has the use of online news sources (Pew Research Center for the People and the Press, 2008). However it is still unclear how much young adults use these new media for personal use or for academic research. A recent study finds use of social media and online news among young adults appears to be declining, while use of these media is growing with 'older consumers' (Benkoil, 2010). Young adults seeking to track election coverage on the Internet tended to rely on 'Old Brand' news organizations and often avoided online news video (Vahlberg, et al., 2008) and polls consistently give television news lower credibility and bias ratings (Pew Research Center. Project for Excellence in Journalism, 2009c).

Student use of new media content for research purposes is likely largely impacted by their research methods ('power browsing') and by faculty recommendations and advice (Head & Eisenberg, 2009). One study argues, whether faculty incorporate new media into their teaching and research 'is determined by personal teaching style and philosophy,' but the greatest use of image and visual materials, including news and other media resources, is in the classroom (Harley, 2007). Librarians may also need to begin to re-think the role of news sources in their information literacy efforts. It is also possible the number of news stories, blogs, and Twitter feeds related to any news event and the repetitious nature of news stories as they rely increasingly on wire services may make news content too overwhelming for students seeking efficiency rather than thoroughness.

Evidence suggests librarians and faculty approach teaching information gathering skills quite differently. Librarians recommend students move from the general to the specific systematically (Head & Eisenberg, 2009); faculty tend to

emphasize use of specific sources and evaluating content for bias over process and would like to see their students read more deeply; likely faculty are less concerned about whether web content is used, if it has been thoroughly vetted (McClure & Clink, 2009). The clarity of the assignment and clear guidelines on sources to be used likely influences which sources students use (Barratt et al., 2009).

As students leverage scholarly e-journal content and use fewer sources for their research, new media sources may be more difficult to locate (Spink & Jansen, 2004a) and be less acceptable to their instructors as reliable sources. Both faculty and students may use sources such as *The Drudge Report*, *Huffington Post*, *Salon*, and *Slate* for their personal information gathering, however polls indicate these sources are generally considered less credible-- thus indicating an inherent conflict in the use of these sources and how they are perceived in the academic and research context (Pew Research Center. Project for Excellence in Journalism, 2009c).

The reliability and credibility of new media in the research framework has not been fully explored but their impact on how news is gathered and distributed is unmistakable. Sports blogs, once considered outside the mainstream are now widely accepted (Weintraub, 2009); the *Pew New Media Index* (http://www.journalism.org/news_index/100) captures the leading commentary of blogs and social media sites focused on news and compares those subjects to that of the mainstream press. Journalists now use blogs in a variety of ways (Cyberjournalist.Net, 2005) and the use of Twitter has greatly impacted the way news is distributed and gathered (Schulte, 2009). While historical news content is clearly connected to the newspaper format, future news sources will be much more diverse, immediate, and changing. In this context, information cycle, timeline, and news event approaches to teaching news content may no longer reflect today's news environment. With 24-hour news cycles, constant updating and commentary on news events on blogs and Twitter, for example, librarians may need to help students better understand how to 'narrow' what news and commentary sources they search and how to search them more effectively, rather than ignoring new media as sources of news content. In light of the 'wealth' of new media forms, librarians may also want to reconsider the research methods and steps they recommend to students. With so much citizen-based journalism and other news commentary in the form of new media, it may no longer be feasible to recommend students systematically move through a research process from general to specific. It is also important to determine whether researchers are seeking access to the new media news and in what context (presentations vs. research papers, for example) and what features (embedded links that work through the proxy server, for example) are needed to make the news video most useful in teaching and research.

Given the changes in the news industry toward increased use of new media to distribute news and to comment or elaborate on news events and the increased use of visual media related to news events, librarians, just as journalism educators have, should seek to incorporate these resources more fully into their information literacy efforts. Further, as librarians incorporate these sources into their course-related instruction, they will need to begin to develop criteria for evaluating and distinguishing between blogs, their commentators, and their

supporters, for example, or how they can be used in specific types of assignments (for example, to support pro/con arguments)—what Stony Brook University School of Journalism calls ‘news literacy’ (<http://www.newsliteracyconference.com/content/>)

Use of Aggregator Databases

Aggregator Databases were once highly prized by librarians and faculty (Oulanov & Pajarillo, 2003). Yet, at Penn State, the use of Pennsylvania titles is now sustaining the use of its Aggregator Databases. Those Aggregator Databases which do not include Pennsylvania content (or make it difficult to search, for example including it as part of the state newswire, for example) are experiencing less use of their news content aside from the ‘Old Brand’ titles. This change has developed over the last five years. While these databases still contain the widest variety of newspapers and greatest name recognition, use of Specialized Databases, specific Historical Databases, and Pennsylvania and New Media content within the Aggregator Databases is growing.

Overall use of Penn State’s news Aggregator Databases is declining. The number of *documents viewed* has changed from 85% (2006) to 61% (2009) in Aggregator Databases. LexisNexis provided 42% of *documents viewed* in 2006 and provides 34% in 2009 [Table 2B: Database Use by Category, 2006-2009, by Documents Viewed]. The number of *searches* in Aggregator Databases has been significantly higher in *ProQuest* since 2006 (86%) as newspaper content is cross-searched with magazine and scholarly journal content. However, use of *Newsbank* is edging up from 2% (2006) to 8% (2009) and now equals the same percentage of searches conducted in *LexisNexis*. On the other hand, while users are searching *ProQuest* significantly more (84%) in 2009 [Table 2A: Database Use by Category, 2006-2009, by Searches] the number of documents viewed (35%) is nearly equal to the documents viewed in *LexisNexis* (34%) and *Newsbank* (31%) in 2009 [Table 2B: Database Use by Category, 2006-2009, by Documents Viewed]. This change implies cross-searching news content with other resources (scholarly journals and magazines) does not necessarily increase the number of documents viewed in newspapers and specific content (‘Old Brands,’ major city, new media, and Pennsylvania, specifically), and this trend likely will influence which database researchers use.

There are also other reasons for this trend. As more news content becomes freely available on the Internet, users can now have many more freely available new resources from which they can ‘hunt and gather what they want, when they want it, use search to comb among destinations..., and share what they find through ...social media’ (Pew Research Center. Project for Excellence in Journalism, 2009a). On the other hand, as *The State of the Media* report finds: many Americans appear to be confused over the source of news. Pew notes many people consider *Google News* and *Yahoo News* the most reliable news sources, not understanding that such aggregators ‘derive much of their content from traditional news organizations’ (Pew Research Center. Project for Excellence in Journalism, 2009c). However, additional factors should be explored, such as the search interface (the use of the Sports category in *PressDisplay*, for example), years of coverage, and completeness of coverage for titles beyond ‘Old Brands.’ Researchers expect online news aggregators to present news more visually to

present both visual and text content simultaneously. In this context, the libraries' Aggregator Databases, while providing more powerful search options, may not allow researchers to dial down their aperture in ways that are meaningful to them.

As the news industry seeks ways to 'harness' its content, Pew has noted the need for news 'online retail malls' and Vahlberg (2008) argues there is a 'large unmet need for a different kind of news site—one that is designed not for news junkies but for inexperienced news consumers.' Generally as users have more options and understand less about who creates news content, they are beginning to lose connection with libraries as news content providers. Possibly libraries need to begin to re-position themselves as a different kind of news source - one for inexperienced news consumers and researchers seeking news content for news-based research.

One professor has argued that 'today, news is increasingly present tense . . . this places an added responsibility on faculty members to help guide their students through the Internet's infinite digital wilderness' (Academe and the decline of news media: Garrick Utley). Libraries and librarians need to also explain their role in the digital wilderness. While the 'wealth' of library news databases is considered a positive development, it has also created a great deal of confusion for student and faculty researchers who are unsure which newspapers are in which databases and whether complete or selected content is available. This may explain why use of news magazines is increasing—by providing news with a week or month's hindsight they simplify and highlight what is or is not important (Barratt et al., 2009). If Head and Eisenberg's (2009) description of student research methods is accurate, then students will gravitate toward known resources that provide the information resources they need quickly and effectively. In this context, libraries may need to find more effective ways to explain how their news resources differ from such sources freely available on the Internet, make them more visible and to do so in ways that are meaningful to students who may not read newspapers and who 'power browse' their way through web and library resources.

It may be unclear to student researchers, aside from the visibility and recognition of 'Old Brands,' how to locate, distinguish between, and use the constant barrage of news content and commentary. Quite simply, students may be 'overwhelmed' by the sheer volume of news sources accessible to them—as one professor notes: 'while the proliferation of information gathering on the internet has some benefits, it also creates an overload of competing factoids and claims whose reliability is hard for the average person to check' (Academe and the decline of news media: Stephanie Coontz).

Today's news consumer expects news to 'find them' and to be delivered to them (Herther, 2009). Yet, many researchers experience many challenges locating a library's online news resources. Libraries have not focused on how to make their news resources visible in today's virtual library or in the Information Commons. As libraries re-design their library website and implement link resolvers they should consider how these same tools can also be used to improve access to news content if implemented carefully. However, news research has a need for slightly different finding tools—ones that help them decide where to search

(such as *NewsCat* <http://www.libraries.psu.edu/psul/researchguides/matbytype/newscat.html>) or that once found, link them directly to a library's news resources, particularly as the media industry turns increasingly to fee-based access to their news content. Such tools need to become part of the freshmen learning curve in the same way a database, such as *ProQuest*, and link resolvers have become.

Libraries have begun to embrace *Google Scholar* by incorporating a search box on their home page and *Google Books* through partnerships with OCLC. However, there is, as yet, little evidence libraries have begun to use *Google News* as an alternative or complement to a library's Aggregator Database. As more research begins to highlight the differences between *LexisNexis* and search engines such as *Google News* (Weaver & Bimber, 2008), librarians should continue to evaluate how these sources meet the needs of libraries in providing access to content their students and faculty need for teaching, research, and learning, not just for reading and personal research needs. Some evidence suggests students will change their search tactics if a categorized database search interface is presented (Kules & Shneiderman, 2007) and use of Sports *category* news content implies a categorized view may work, at least, with some news content.

In light of the changing news environment and higher cost of all news resources, informal evidence suggests libraries have begun to make significant changes in their news collections cancelling print and microfilm subscriptions and many libraries now rely almost entirely on Aggregator Databases (i.e., *LexisNexis*, *Newsbank*, *EBSCO*, *ProQuest* or *Factiva*) to provide access to news content. Clearly libraries have been forced to balance these decisions against the growing costs of all library resources. Given the lack of data and insights into how faculty and students use news content specifically, libraries appear to be making these decisions based on wide-spread evidence of a dying print news industry; evidence of declining newspaper readership; and what appears to be widespread availability of news content freely available on the Internet. They may also believe that their news databases are heavily used given the wealth of resources they make accessible.

CONCLUSION

At present, it is unclear what drives the use of library or web-based news sources for research and teaching rather than personal use. In addition, it appears faculty, students, and librarians use and perceive the role of news databases quite differently in teaching, learning, and research. Thus, any library seeking to better understand how their online news sources are used should study their vendor database use statistics and begin to monitor the actual news and information seeking behaviour of their users at the same time it is considering overall trends in use of library and news resources, generally.

As more news resources become freely available and searchable, libraries are no longer the sole gateway or the most visible provider of news content, yet a library's wealth of news database provides distinct differences from freely available Internet based search engines and news sites. Library web pages,

library instruction, and finding tools continue to focus on newspapers. However, as news content becomes disaggregated from the newspaper format, libraries may be even less associated with news content as many students do not equate news with a newspaper and may not associate web-based news with libraries. As they have with scholarly e-journals and databases, libraries should seek to understand how and why researchers do (or do not) use the library's news resources and find ways to make them more visible, easier to locate, and to more effectively articulate how these library news resources differ from *Yahoo News* or *Google News*, keeping in mind today's faculty and student researchers increasingly are focused on efficiency and effectiveness as they seek information to support their teaching, research and learning. Librarians need to better understand how newspapers, as compared with news content, can be and is used in leisure reading, teaching, and research within the academic and research framework—and whether these collections are important in establishing the library as a virtual place for news content.

Generally it appears the use of library-provided news resources will lag general trends, but use of library news resources is already declining as they face a wider range of competition from freely available news sources, more electronic sources such as e-journals and magazines, and concern about the credibility of new media forms, such as blogs and television. Over the long term, the use of library-provided news resources will mirror larger news readership trends and libraries will need to position themselves to ensure researchers and students continue to perceive and understand them to be a resource for news content in every format. There are still many changes yet to take place in the news industry and in libraries. Librarians must stay abreast of these changes and continue to seek ways to position their resources for tomorrow's research needs. That is the library's niche today and tomorrow.

**Table 1: "Old Brand" Use Compared - Aggregator Databases, 2009
by Documents Viewed**

Brand	Total	% of Total	Newsbank	% of Newsbank	Proquest	% of Proquest	LexisNexis*	% of LexisNexis
New York Times	63,598	25%	9,742	13%	39,823	24%	14,033	90%
Washington Post	24,241	10%	4,525	6%	18,210	11%	1,506	10%
Los Angeles Times	15,983	6%	667	1%	15,316	9%	-	0%
Chicago Tribune	14,084	6%	2,429	3%	11,655	7%	-	0%
USA Today	10,678	4%	2,161	3%	8,460	5%	57	0%
Total 'Old Brands'	128,584	50%	19,524	26%	93,464	57%	15,596	100%
Wall Street Journal	2,839	1%	225	0%	2,609	2%	5	0%
Other Major US Cities	81,776	32%	21,522	29%	60,254	37%	-	0%
Centre Daily Times	12,088	5%	12,088	16%	-	0%	-	0%
Philadelphia Inquirer	6,905	3%	5,954	8%	951	1%	-	0%
Pittsburgh Post Gazette	11,452	4%	4,299	6%	7,153	4%	-	0%
Other PA newspapers	11,175	4%	11,175	15%	-	0%	-	0%
Total PA Newspapers	41,620	16%	33,516	45%	8,104	5%	-	0%
TOTALS	254,819	100%	74,787	100%	164,431	100%	15,601	100%

*LexisNexis numbers may be underrepresented due to publications being accessed through L/N categories such as Major World Publications

Table 2A: Database Use by Category, 2006 – 2009 by Searches

SEARCHES (Combination)	2006	% 2006	2007	% 2007	2008	% 2008	2009	% 2009	Total	% Total
Searches (Aggregator Databases)	1,547,148	53%	1,447,491	54%	1,510,725	53%	1,468,706	53%	5,974,070	53%
Searches (Specialized)	1,271,480	44%	1,152,970	43%	1,257,296	44%	1,218,508	44%	4,900,254	44%
Searches (Historical)	75,497	3%	64,826	2%	85,128	3%	73,166	3%	298,617	3%
Total by Year	2,894,125	100%	2,665,287	100%	2,853,149	100%	2,760,380	100%	11,172,941	100%
Vendor (HISTORICAL) Searches										
Gale: 19th Century British Newspapers	-	-	-	-	2,403	3%	1,165	2%	3,568	1%
Gale: British Newspapers	-	-	-	-	1,598	2%	4,235	6%	5,833	2%
Gale: 19th Century US Newspapers	-	-	-	-	10,438	12%	7,696	11%	18,134	6%
Gale: The Times Digital Archive	-	-	-	-	7,839	9%	6,656	9%	14,495	5%
Proquest Historical (NY Times)	68,813	91%	50,090	77%	49,306	58%	38,132	52%	206,341	69%
Newsbank: Americas Historical Newspapers	6,684	9%	14,736	23%	13,544	16%	15,282	21%	50,246	17%
Total by Year	75,497	100%	64,826	100%	85,128	100%	73,166	100%	298,617	100%
Vendor (SPECIALIZED) Searches										
Eastview	1,205	0%	693	0%	657	0%	2,443	0%	4,998	0%
Ethnic NewsWatch & Black Newspapers	1,270,275	100%	1,152,277	100%	1,253,949	100%	1,207,054	99%	4,883,555	100%
World News Connection	-	-	-	-	2,690	0%	9,011	1%	11,701	0%
Total by Year	1,271,480	100%	1,152,970	100%	1,257,296	100%	1,218,508	100%	4,900,254	100%
Vendor (AGGREGATOR DATABASES) Searches										
LexisNexis	173,197	11%	166,622	12%	138,921	9%	112,996	8%	591,736	10%
Newsbank (Full Text Newspapers)	36,404	2%	39,120	3%	68,039	5%	124,650	8%	268,213	4%
Proquest (PQ Newspapers)	1,337,547	86%	1,241,749	86%	1,303,765	86%	1,231,060	84%	5,114,121	86%
Total by Year	1,547,148	100%	1,447,491	100%	1,510,725	100%	1,468,706	100%	5,974,070	100%

Table 2B: Database Use by Category, 2006 – 2009 by Documents Viewed

DOCS VIEWED (Combination)	2006	% 2006	2007	% 2007	2008	% 2008	2009	% 2009	Total	% Total
Docs Viewed (Aggregator Databases)	618,883	85%	549,364	77%	614,807	72%	464,440	61%	2,247,494	74%
Docs Viewed (Specialized)	32,494	4%	39,290	6%	74,828	9%	160,452	21%	307,064	10%
Docs Viewed (Historical)	77,639	11%	121,861	17%	163,952	19%	135,287	18%	498,739	16%
Total by Year	729,016	100%	710,515	100%	853,587	100%	760,179	100%	3,053,297	100%
Vendor (HISTORICAL) Docs Viewed										
Gale: 19th Century British Newspapers	-	-	-	-	12,240	7%	2,184	2%	14,424	3%
Gale: British Newspapers	-	-	-	-	4,643	3%	8,361	6%	13,004	3%
Gale: 19th Century US Newspapers	-	-	-	-	12,183	7%	10,668	8%	22,851	5%
Gale: The Times Digital Archive	-	-	-	-	5,231	3%	4,851	4%	10,082	2%
Proquest Historical (NY Times)	56,680	73%	52,230	43%	83,940	51%	55,188	41%	248,038	50%
Newsbank: Americas Historical Newspapers	20,959	27%	69,631	57%	45,715	28%	54,035	40%	190,340	38%
Total by Year	77,639	100%	121,861	100%	163,952	100%	135,287	100%	498,739	100%
Vendor (SPECIALIZED) Docs Viewed										
Eastview	1,311	4%	749	2%	30,927	41%	117,891	73%	150,878	49%
Ethnic News Watch & Black Newspapers	31,183	96%	38,541	98%	43,796	59%	42,100	26%	155,620	51%
World News Connection	-	-	-	-	105	0%	461	0%	566	0%
Total by Year	32,494	100%	39,290	100%	74,828	100%	160,452	100%	307,064	100%
Vendor (AGGREGATOR DATABASES) Docs Viewed										
LexisNexis	262,753	42%	224,458	41%	312,033	51%	155,689	34%	954,933	42%
Newsbank (Full Text Newspapers)	55,240	9%	50,086	9%	103,623	17%	144,320	31%	353,269	16%
Proquest (PQ Newspapers)	300,890	49%	274,820	50%	199,151	32%	164,431	35%	939,292	42%
Total by Year	618,883	100%	549,364	100%	614,807	100%	464,440	100%	2,247,494	100%

Table 3: Ratio of Searches to Documents Viewed, 2006 – 2009

Ratio of Searches to Documents Viewed	2006	2007	2008	2009
Aggregator Databases (Searches/Doc Viewed)	2	3	2	3
Specialized (Searches/Doc Viewed)	39	29	17	8
Historical (Searches/Doc Viewed)	1	1	1	1
TOTAL (Searches/Doc Viewed)	4	4	3	4
Ratio of Searches to Docs Viewed (Historical)				
Gale: 19th Century British Newspapers	-	-	0	1
Gale: British Newspapers	-	-	0	1
Gale: 19th Century US Newspapers	-	-	1	1
Gale: The Times Digital Archive	-	-	1	1
Proquest Historical (NY Times)	1	1	1	1
Newsbank: Americas Historical Newspapers	0	0	0	0
Total by Year	1	1	1	1
Ratio of Searches to Docs Viewed (Specialized)				
Eastview	1	1	0	0
Ethnic NewsWatch & Black Newspapers	41	30	29	29
World News Connection	-	-	26	20
Total by Year	39	29	17	8
Ratio of Searches to Docs Viewed (Aggregator Databases)				
LexisNexis	1	1	0	1
Newsbank (Full Text Newspapers)	1	1	1	1
Proquest (PQ Newspapers)	4	5	7	7
Total by Year	2	3	2	3

Table 4: Newspaper Requests from Inter-Library Loan (ILL), 2004 – 2008

Year Requested	2004	2005	2006	2007	2008	Total
Pre 1924	148	128	125	95	109	605
1925-1949	74	52	19	22	41	208
1950-1974	18	33	14	29	33	127
1975-1999	19	53	13	11	17	113
(From preceding page)						
2000 - Current	4	11	14	28	12	69
Unidentified Year	16	25	16	5	9	71
Total	279	302	201	190	221	1,193
Status of Requestor						
Faculty	120	157	96	67	80	520
Graduate	50	96	83	83	102	414
Law Student	-	-	-	-	1	1
Retiree	-	-	-	-	2	2
Scholar	4	9	5	4	-	22
Staff	76	23	12	26	17	154
Undergraduate	29	17	5	10	19	80
Total	279	302	201	190	221	1,193

(Contd... next page)

(From preceding page)

College	2004	2005	2006	2007	2008	Total
Agricultural Sciences	-	3	1	2	1	7
Arts and Architecture	14	28	15	1	13	71
Smeal College of Business	-	-	5	-	-	5
Communications	10	13	13	15	16	67
Earth and Mineral Sciences	-	6	6	5	-	17
Education	12	6	2	1	12	33
Engineering	2	11	4	2	5	24
Health and Human Development	23	-	16	28	9	76
Information Sciences & Technology	-	-	-	-	-	-
Liberal Arts	156	133	83	100	128	600
Medicine & Nursing	-	-	-	-	-	-
Graduate School	-	-	-	-	7	7
Schreyer Honors College	3	-	-	-	-	3
Other (Campuses, Administration, & Libraries)	59	102	56	36	30	283
Total	279	302	201	190	221	1,193

Table 5: Percent of Historical News Content Use Compared to Aggregator Databases, 2006 – 2009

	2006	2006	2007	2007	2008	2008	2009	2009
Aggregator Databases	618,883	89%	549,364	82%	614,807	79%	464,440	77%
Gale	-	0%	-	0%	34,297	4%	26,064	4%
Vanderbilt	129	0%	54	0%	985	0%	478	0%
America's Historical Newspapers (READEX)	20,959	3%	69,631	10%	45,715	6%	54,035	9%
Proquest The New York Times Historical	56,680	8%	52,230	8%	83,940	11%	55,188	9%
TOTALS	696,651	100%	671,279	100%	779,744	100%	600,205	100%

Table 6: Use of International Sources: Newsbank, 2009 by Documents Viewed

	Newsbank 2009	% of total
Africa/Middle East	1,850	1.31%
Asia	3,311	2.34%
Australia	897	0.63%
Europe	2,062	1.46%
North/Latin America	981	0.69%
PA	33,586	23.77%
USA	98,608	69.79%
Total	141,295	100.00%

Table 7: Newspaper Use Online and In Room, 2009 by Issues Read

	USA	%USA	PA	% PA	Business	% Business	Asia	% Asia	Sport	% Sport	Europe	% Europe	Latin/ North America	% Latin/ North America	Middle East/ Africa	%Middle East/ Africa	Total
Press Display	3,731	40%	370	4%	603	6%	1,167	12%	256	3%	2,178	23%	779	8%	291	3%	9,376
In Room*	612	21%	381	13%	496	17%	790	27%	-	-	437	15%	100	3%	76	3%	2,893
Total	4,343	35%	751	6%	1,099	9%	1,957	16%	256	2%	2,615	21%	879	7%	367	3%	12,269

Table 8: Use of New Media, 2004 - 2009

	2004	2005	2006	2007	2008	2009
Web	-	-	1	330	1,417	2,613
Blogs	-	-	-	72	294	652
Video	232	156	129	56	1,176	676
Total	232	156	130	458	2,887	3,941

Table 9: Use of New Media by Vendor, 2004 - 2009

	2004	2005	2006	2007	2008	2009
Vanderbilt	232	156	129	54	985	478
Newsbank	-	-	18	12	442	2,945
LexisNexis	31,207	7,918	5,818	5,767	11,278	8,867
Total	31,439	8,074	5,965	5,833	12,705	12,290

REFERENCES

1. Academe and the decline of news media: Garrick Utley. (2009, November). *Chronicle of Higher Education*.
2. Academe and the decline of news media: Stephanie Coontz. (2009, November). *Chronicle of Higher Education*.
3. Barratt, C. C., et al. (2009). Collaboration is key: Librarians and composition instructors analyze student research and writing. *portal: Libraries and the Academy* 9(1), 37-56.
4. Becker, C. H. (2009). Student values and research: Are millennials really changing the future of reference and research? *Journal of Library Administration*, 49(4), 341-365.
5. Benkoil, D. (2010). Young online newspaper audience dropping: Older consumers adopting social media. *E-Media tidbits (Poynter Online)*, Jan. 22, 2010. Retrieved from <http://www.poynter.org/column.asp?id=31&aid=176279>
6. Budd, J. M. & Christensen, C. (2003). Social sciences literature and electronic information. *portal: Libraries and the Academy*, 3(4), 643-651.
7. Bui, C. L. (2009). Examining the new gatekeepers: News portals' inclusion and ranking of media and events. Paper presented International Communication Association.
8. Carlson, J. (2006). An examination of undergraduate student citation behavior. *The Journal of Academic Librarianship*, 32(1), 14-22.
9. Caudle, D. M. & Schmitz, C. M. (2007). Web access to electronic journals and databases in ARL libraries. *Journal of Web Librarianship*, 1(1), 3-26.
10. Cyberjournalist.Net. (2005). *How journalists use blogs*, Feb. 2, 2010. Retrieved from <http://www.cyberjournalist.net/news/002821.php>
11. Darst, David H. (1991). Spanish-language newspapers in the classroom. *Hispania*, 74(1), 202-204.
12. Dezsö, Z., et al. (2006). Dynamics of information access on the web. *Physical Review*, E 73, 066132-1-066132-6.
13. Freeland, M. & Bailey, M. (2008). Print newspapers: Are they still being used in academic and research libraries? *Serials Librarian*, 55(1), 210-226.
14. Fuller, K., et al. (2009). Making unmediated access to E-resources a reality: Creating a usable ERM interface. *Reference & User Services Quarterly*, 48(3), 287-301.
15. Harley, D. (2007). Why study users? An environmental scan of use and users of digital resources in humanities and social sciences undergraduate education. *First Monday*, 12(1). Retrieved from Available at http://firstmonday.org/issues/issue12_1/harley/index.html
16. Head, A. J. & Eisenberg, M. B. (2009). *How college students seek information in the digital age: Lessons learned* (Project Information Literacy. Progress Report). Seattle, WA: University of Washington, The Information School. Retrieved from http://projectinfolit.org/pdfs/PIL_Fall2009_Year1Report_12_2009.pdf

17. Herther, N. K. (2009). 'If the news is that important, it will find me': The future of journalism, newspaper, and finding information. *Searcher*, 17(4), 22-32.
18. *Information behaviour of the researcher of the future: A CIBER briefing paper* (2008). University College London, JISC. Retrieved from <http://www.bl.uk/news/pdf/googlegen.pdf>
19. Kriebel, L. & Lapham, L. (2008). Transition to electronic resources in undergraduate social science research: A study of honours theses bibliographies, 1999-2005. *College & Research Libraries*, 69(3), 268-283.
20. Kules, B. & Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management*, 44, 463-484.
21. Kyriolidou, M. & Bland, L. (2009). *ARL Statistics 2007-08*. Washington, D.C.: Association for Research Libraries. Retrieved from: <http://www.arl.org/stats/annualsurveys/arlstats/arlstats08.shtml>
22. Leiding, R. (2005). Using citation checking of undergraduate honours thesis bibliographies to evaluate library collections. *College & Research Libraries*, 66(5), 417-429.
23. McCargar, V. (2005). Following the trail of the disappearing data. *The Seybold Report*, 4(21), 7-14. Retrieved from <http://www.loc.gov/standards/premis/TSR-0209.pdf>
24. McClure, R., & Clink, K. (2009). How do you know that? An investigation of student research practices in the digital age. *portal: Libraries and the Academy*, 9(1), 115-132.
25. Oulanov, A. & Pajarillo, E. J. Y. (2003). Academic librarians' perception of Lexis-Nexis. *Electronic Library*, 21(2), 123-129.
26. Pew Research Center for the People and the Press. (2008). *Key new audiences now blend online and traditional sources: Audience segments in a changing news environment*. Washington, D.C.: The Center. Retrieved from <http://people-press.org/report/444/news-media>
27. Pew Research Center for the People and the Press. (2009). *Many would shrug if their local newspaper closed*. Washington, D.C.: The Center. Retrieved from <http://people-press.org/report/497/many-would-shrug-if-local-newspaper-closed>
28. Pew Research Center for the People and the Press. (2010). *How news happens: A study of the news ecosystem of one American city*. Washington, D.C.: The Center. Retrieved from http://www.journalism.org/analysis_report/how_news_happens
29. Pew Research Center. Project for Excellence in Journalism. (2009a). *State of the news media 2009* (Annual). Washington, D.C.: The Center. Retrieved from <http://www.stateofthedia.org/2009/index.htm>
30. Pew Research Center. Project for Excellence in Journalism. (2009b). *The state of the news media 2009: Overview, major trends*. Washington, D.C.: The Center. Retrieved from http://www.stateofthedia.org/2009/narrative_overview_intro.php?media=1

31. Pew Research Center. Project for Excellence in Journalism. (2009c). *The state of the news media 2009: Overview, public attitudes*. Washington, D.C.: The Center.
32. Quint, B. (2009). Where have all the archives gone? Newspaper archive aggregators face the challenge of all-digital, no-paper publications. *Information Today*, 26(7), 1-38.
33. Rettig, J., et al. (2004). *The many uses of newspapers*. Richmond, VA: University of Virginia, Boatwright Library. Retrieved from <http://oncampus.richmond.edu/.../The%20Uses%20of%20Newspapers.doc>
34. Schulte, B. (2009). The distribution revolution. *American Journalism Review*, 31(5), 22-25.
35. Spink, A., & Jansen, B. J. (2004a). Multimedia searching. *Web search: Public searching of the web* (pp. 161-177). Boston: Kluwer Academic Pub.
36. Spink, A., & Jansen, B. J. (2004b). *Web search: Public searching of the web*. Boston: Kluwer Academic Pub.
37. Tapscott, D. (2008). Net genres relate to news in new ways. *Nieman Reports*, 62(4), 18-19.
38. Tenopir, C. (2003). *Use and users of electronic library resources: An overview and analysis of recent research studies*. Washington, D.C.: Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub120/pub120.pdf>
39. Thorson, E. (2008). Changing patterns of news consumption and participation. *Information, Communication & Society*, 11(4), 473-489.
40. Vahlberg, V., et al. (2008). *From 'too much' to 'just right': Engaging millennials in election news on the web*. Northwestern University, Media Management Center. Retrieved from <http://www.mediamanagementcenter.org/research/youthelection.pdf>
41. Vahlberg, V. (2008). What young people don't like about the web--and news on it. *Nieman Reports*, 62(4), 30-32.
42. Weaver, D. A. & Bimber, B. (2008). Finding news stories: A comparison of searches using LexisNexis and Google News. *Journalism & Mass Communication Quarterly*, 85(3), 515-530.
43. Weintraub, R. (2009). Into the fold. *Columbia Journalism Review*, 48(2), 10.
44. Wiggins, R. W. (2002). Yesterday's headlines. *Library Journal* (1976), 26-27.
45. World Association of Newspapers. *Youth media DNA: Decoding youth media and information consumers*, 2/2/2010, from <http://www.wanpress.org/nic/articles.php?id=681>
46. Xie, I. & Wolfram, D. (2009). A longitudinal study of database usage within a general audience digital library. *Journal of Digital Information*, 10(4) Retrieved from <https://journals.tdl.org/jodi/article/viewArticle/304>

BUILDING A 24x7 NEWSPAPER DIGITAL LIBRARY AND ARCHIVES

Case Study of DNA – Daily News and Analysis

Anita Pujari

ABSTRACT

The year 1779 marked the birth of The Hickey's Bengal Gazette or the Calcutta General Advertiser, the first printed English newspaper in the Indian subcontinent in Calcutta. The year 2005 saw the birth of DNA- Daily News and Analysis in Mumbai, India. These two newspapers are centuries apart not just in their look and style, content and format, production and distribution but also in their end-user the 'newspaper reader'. However, one thing is common and that is – the intent to provide information and chronicle the times we live in through a mass media tool, namely, a newspaper.

This paper will not delve into history but in the backdrop of the evolution of a newspaper, put forth the theory and practice of an evolving Newspaper Library. Through this paper I would like to share my experience in setting up a 24x7 Digital Library and Digital Archives in a new born newspaper. The concept of a single window access to the complete Digital Archives of the newspaper as seen in its final electronic format or the E-paper and the components both published and unpublished as in stories, photographs, illustrations and cartoons is discussed. The strategy adopted to expand this single window 24x7 access includes all the resources and tools which are essential and editorial support that a newspaper library needs to provide. Touching upon these essential tools the paper elaborates the method of building and fine tuning them to meet the changing needs of newspaper creators today. The library professionals in a newspaper library must cater to the Marketing, Sales, Circulation and Production departments as much as they do for the Editorial department.

How does one measure the performance and efficacy of a service department? The paper shares the methodology put in place to do so. How the value addition to the organization is computed and projected to the management is elaborated upon.

This paper will also touch upon the successful initiative of monetizing the Digital Archives at DNA. Creation of a Syndication business for the company and converting the newspaper Library and Archives named Research N Archives department in DNA from a cost centre to a profit centre is also discussed.

In conclusion this case-study brings out the potential of newspaper library professionals to play an important role in adding value to a newspaper company. We are professionally equipped to support the commercial aspect of today's market driven newspapers as well as give timely editorial support to keep the edge of quality content. Oft debated and as often agreed upon "Content is King" and this is the differentiator that 1.9 billion daily newspaper readers around the

world seek. India accounts for 107 million newspaper sales every day making her the biggest market for newspapers in the world. We have a great future!

Keywords: 24x7 Digital Library, Digital Archives, Newspaper Library, Content Management, Stories, Photographs, Syndication, E-Paper, Valuation, Performance Analysis, DNA-Daily News and Analysis

INTRODUCTION

"Despite the endless predictions about the death of newspapers, they actually continue to grow, at least on a global scale," said Timothy Balding, co-CEO of WAN-IFRA as he presented the World Press Trends Survey at the World Association of Newspapers WAN-IFRA conference at Hyderabad, India in December 2009. The WAN-IFRA survey showed that newspaper circulation grew, on a global scale, by 1.3 percent in 2008, the last full year for which data exists, and almost 9 percent over five years.

Globally, 1.9 billion people choose to read a newspaper every day, or 34 percent of the world population, while 24 percent use the Internet. The biggest newspaper market in the world is India, with 107 million daily sales. India, China and Japan account for more than 60 percent of world sales, with the USA taking 14 percent.

In a 182-billion dollar press advertising industry, digital revenues of newspapers accounted for less than 6 billion dollars in 2008 and as forecast by PricewaterhouseCoopers (PwC) it would grow to no more than 8.4 billion dollars by 2013. "These PwC forecasts, similar to those made by Zenith Optimedia and others, demonstrate quite simply that at no time soon will digital advertising revenues come close to achieving the sort of revenues required, by many, to compensate for falling print revenue," said Timothy Balding, co-CEO of WAN-IFRA. "So that answer will have to be found elsewhere ... Should these forecasts come close to being true, new business models will have to be invented."

As newspapers continue to evolve meeting new challenges with new business models the Newspaper Library Services and Access Methods need to continuously evolve and keep pace with new developments in order to remain relevant.

This paper is a case study of a newspaper born in the digital age and how its Newspaper Digital Library Services was built to meet the needs of the present day newspaper business.

INTRODUCTION TO DNA – DAILY NEWS and ANALYSIS

DNA – Daily News and Analysis launched its first edition in Mumbai on July 30, 2005. Four more editions have been launched since, from Pune, Ahmedabad, Jaipur and Bangalore. The latest Indian Readership Survey (IRS 2009 R2) pegged DNA as the 7th largest English newspaper in India.

DNA is owned by Diligent Media Corporation Ltd., a joint venture between media industry majors – the Bhaskar Group (publishers of Dainik Bhaskar and Divya Bhaskar, My FM radio and more) and the Essel Group (Zee Television network, Dish TV, Essel world and more)

INTRODUCTION TO DNA – RESEARCH N ARCHIVES: THE CASE STUDY

DNA broke many new grounds with its launch. Amongst them was its exemplary vision of planning a Digital Library and Archive before the first edition rolled out on July 30, 2005.

I joined the organization and had about 2 months to plan and execute my vision which I named Research N Archives (RNA)

The Concept of RNA

“Knowledge centre par excellence adding value across the organization”. Not an easy vision to actualize especially as one had to begin from scratch.

- First step was identifying the right editorial supports and procuring them - Wire Services, Photo Databases, News Archives, Statistical Info, Corporate and Industry Databases etc
- Second step was setting up the Content Management System (CMS) which would capture, assimilate, organize, index and retrieve the DNA content as it rolled out - both published and unpublished. Namely the DNA Digital Archives
- The database architecture and workflow of the Digital Archives Management System (DAMS) was conceived and developed factoring easy integration and scalability of new Editions
- RNA Archival Operations Network is run from Mumbai, Ahmedabad, Pune, Jaipur and Bangalore. Over 1000 updates are added in the Archives everyday
- RNA Mumbai is the central hub, managing the Archives, Quality Check and Audit, Usage monitoring and analytics and new developments.
- RNA Online Library System is the front-end.
- It is a web based single window 24x7 access across all DNA editions in India. The Digital Archives, Information resources, Research and Services are provided through this single sign on window.
- Digital rights management is inbuilt setting accesses as justified by the user's job profile. This serves to protect the Archives, Databases and Wire accesses from being misused without hampering the cause of a single sign on facility
- An interactive tool “Ask RNA” welcomes and delivers on all posted research/info/access queries within minutes. Constant monitoring and delivery by the team across editions is provided.

Strategic development of RNA

Having set into process the primary deliverables of Digital Archives Management, Research and Information services, RNA took on a new challenge in January 2006. A business initiative to monetize the Digital Archives of DNA through Content Syndication was proposed. The proposed business plan would generate additional revenue for the company as well as promote the DNA Brand to new geographies/readers beyond the limits of print circulation. This dual objective of additional revenue and brand propagation was encouraged by the Management and a Content Syndication division was set up within RNA. Group publications Dainik Bhaskar (Hindi newspaper) and Divya Bhaskar (Gujarati newspaper) were also added to the business initiative and a website www.3dsyndication.com took shape. Launched in May 2006 the website is a one-stop-shop for multilingual news content (English, Hindi and Gujarati), Photographs, Cartoons, Caricatures and Illustrations.

In June 2007 the DNA E-Paper moved to a new technology platform and RNA was bestowed with the responsibility of managing it. Subsequently four DNA editions were launched with their E-Papers being added to the global offering <http://epaper.dnaindia.com> from day one. A revenue model was put in place for Archives on Subscription and Online Advertisements

INFORMATION MANAGEMENT and INFORMATION SERVICES

RNA provides essential research support to DNA Editorial and Ad Sales on demand as well as proactively through various services, info-capsules and info alerts. These services are offered through a single window 24x7, online facility on the Intranet.

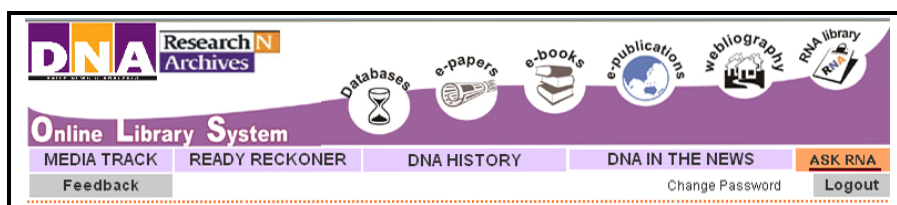


Figure 1: Research N Archives Online Library System – Partial view

Listed below are the key services offered on the **Research N Archives Online Library System**.

- Databases – Links to all subscribed data resources, wire services, imagery. All access/details are available on a click. Digital rights management is inbuilt hence only those permitted can access this section
- E-Papers – links to DNA E-Paper Archives, subscribed E-papers and other. The complete archives of all DNA editions is available on intranet through this system
- E-Books – links to e-books, e-encyclopedias, e-thesaurus, e-dictionaries which could add value

- E-Publications – links to important national and international newspapers and magazine sites
- Webliography - Researched and Recommended websites indexed by subject. Since Google usage is inevitable this is a service that extends our expertise in search-retrieval with oft referred
- RNA Library – lists Books, newspaper and magazines at DNA Edition locations
- Media Track – Daily Current Awareness Services compiling the news and developments in the media industry including Print, Online, Mobile, Television, Radio, Advertising and People movement.
- Ready Reckoners - Quick Reference Info-capsules compiled by RNA. These are proactive research compilations which provide reference and add value as info boxes in the newspaper. Some examples - Quotable Quotes, Chronology of Events like Bomb Blasts, Natural Disasters, Timelines of issues say Kasab trial, Satyam scam, Statistics on Swine-flu, Events/Festivals 2010
- DNA History – the growth of DNA, the changes in the newspaper, new business developments etc are archived by RNA for posterity. DNA Timeline, DNA Newsletter are end products
- DNA in the News – news articles on DNA published in various sources
- ASK RNA – interactive tool for posting research queries

Wire Services Management – selection, liaison, licensing, renewals, billing and payments is a key function of RNA. The responsibility of DNA editorial and subscriptions budget for all the editions running in to a couple of millions rests with RNA. Cost optimization and usage analysis are ongoing activities on this front.

RNA maintains a small but sterling library facility of newspapers, magazines and books which can be availed by all employees.

CONTENT MANAGEMENT - BUILDING THE DNA DIGITAL ARCHIVES

DNA is a born digital newspaper. The Digital Archival System essentially assimilates all the published and unpublished content generated everyday and creates a database of these objects with relevant metadata which would serve to search and retrieve for instant reference and re-use.

The Archival Assets or DNA Intellectual Property comprise the following

- Page Archives
- Story Archives
- Image Archives - Photo Archives, Illustrations, Cartoons and Graphics

The access rights to the Digital Archives are primarily for Editorial and the access is provided on RNA ONLINE LIBRARY SYSTEM. See Fig 2.

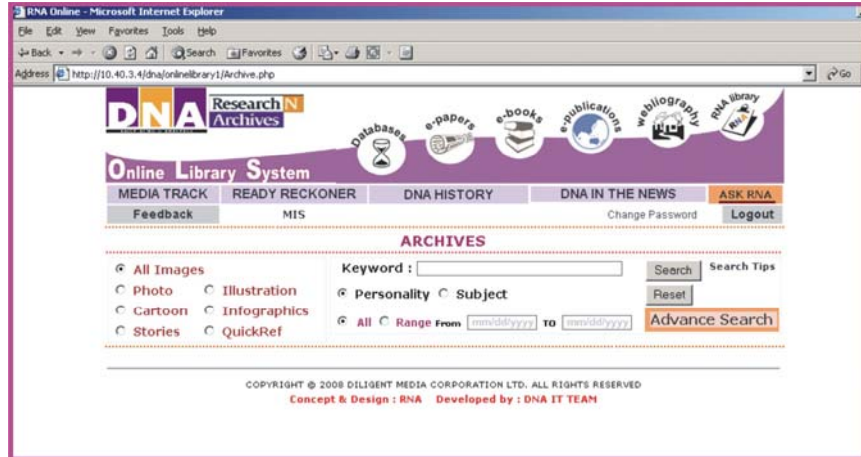


Figure 2: Research N Archives Online Library System – Full view

The Integrated Editorial Management System (IEMS) is the pre-production platform which generates the daily newspaper. Production of DNA is a completely digital process. The Digital Asset Management System (DAMS) draws from IEMS to create the Digital Archives. The Content Management System (CMS) for Syndication draws from both IEMS and DAMS. See Figure 3.

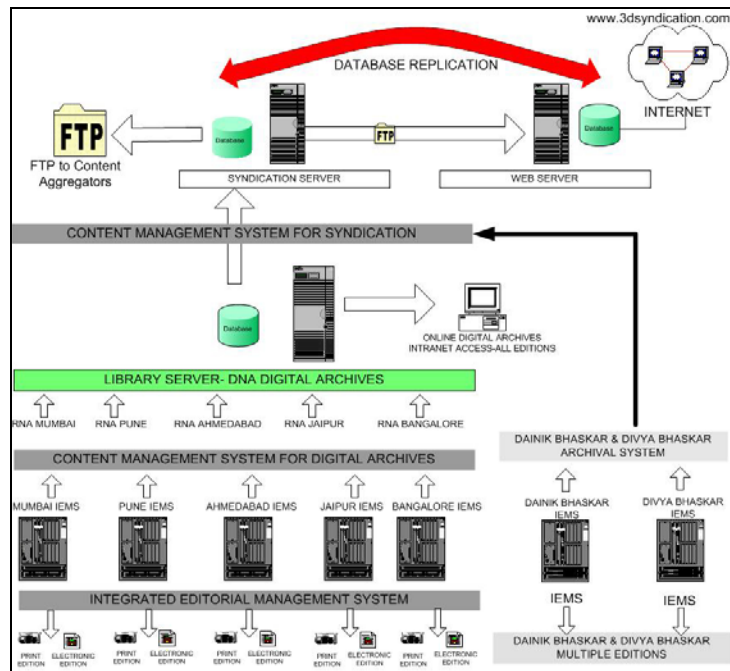


Figure 3: Digital Network and Content Flow

DATABASE AND SOFTWARE TECHNOLOGY

Technology and Platform Used for Archives

- Integrated Development Environment: - **Dreamweaver MX 2004**
- Coding Language: - **PHP**
- Scripting Language: - **Java Script**
- Database: - **SQL Server 2000**

Technology and Platform Used for RNA Online Library System

- Integrated Development Environment: - **Dreamweaver MX 2004**
- Coding Language: - **PHP**
- Scripting Language: - **Java Script**
- Database: - **SQL Server 2000**

Technology and Platform Used for Syndication website

- Integrated Development Environment (i.e. Framework): **.NET (Visual Studio 2005 Version 2.0)**
- Backend Language: **-VB (Visual Basic)**
- Scripting Language: **-Java Script**
- Database: - **SQL Server 2005(Global), SQL Server 2000(Local)**

Standards

Supports industry formats and standards – JPEG, HTML, XML, NEWSML and NITF

IPTC taxonomy with local variations

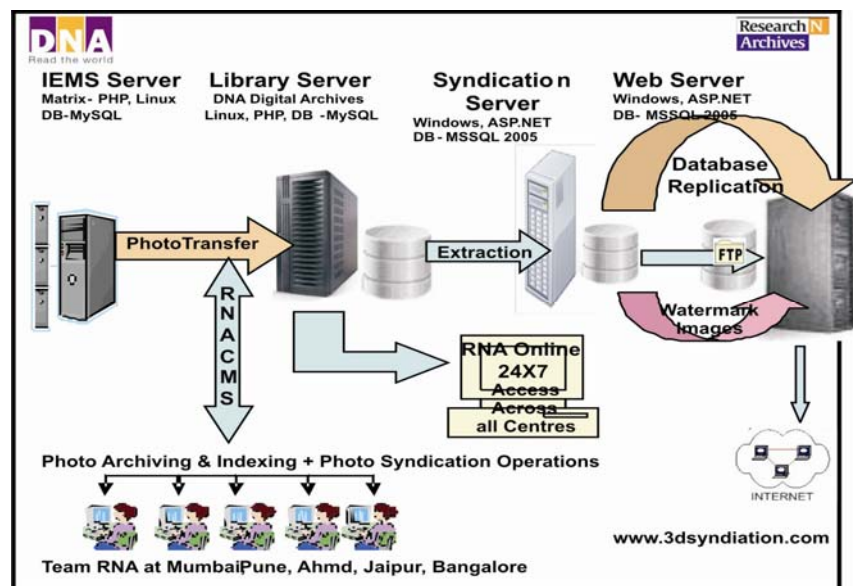


Figure 4: Images Flow-Chart To Archives and Syndication

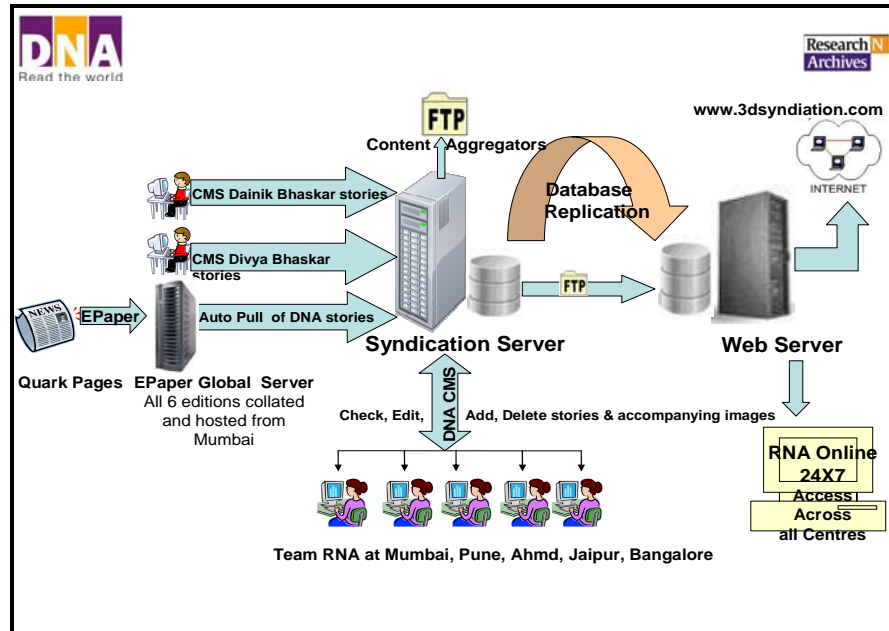


Figure 5: Stories Flow-Chart To Archives and Syndication

CONTENT SYNDICATION – BUSINESS INITIATIVE

This initiative by RNA has served the dual purpose of generating revenue from the Digital Archives and extending the brand reach of DNA to new geographies, platforms and businesses. All activities relating to the business are handled by RNA.

- Syndication operations. The content Management System for Syndication is integrated with the Digital Archives System as shown in Figures 3, 4 and 5.
- Website management www.3dsyndication.com
- Sales – devise strategy to leverage content. Identify partners/customers. Negotiate contracts
- Client Account Management and Customer Servicing
- Marketing – Brand recognition, higher visibility, wider reach
- Legal and Finance

E-PAPER MANAGEMENT

RNA manages the front-end of DNA E-Paper all editions hosted at <http://epaper.dnaindia.com>

It was developed with a revenue model. Access to issues of the past 7 days is on a click. One has to register for access to free archives of the past year. Archives

beyond a year are paid subscription. The second revenue is from advertisements on the site. RNA manages

- E-paper development and site management
- Subscriptions management
- Ad space revenue generation
- Customer services
- Marketing. daily newsletter
- Tracking and analysis

MARKETING RNA

The Sales and Marketing of Syndication and E-Paper apart, RNA aggressively markets itself to the internal user in the organization. This is essential to optimize usage and add value as envisioned.

RNA orientation programme for all newcomers and re-orientation programmes is a means of one to one marketing. RNA ALERTS is e-mailer marketing on the lines of an SDI service. On RNA ONLINE the intranet system we have RNA COMMUNIQUE on the Homepage. This is used to give practical tips, highlight utilities, alert on new additions to the Archives or Ready Reckoner compilations and so on.

Figure 6: RNA Marketing tool

REPORTS and ANALYSIS

RNA has a set system of monitoring and analysis. This is done month on month and compiled as RNA Performance Updates for the Management. It helps project the function, and builds on its brand value.

The numbers are analyzed and these monthly insights are used to fine-tune the growth strategy to achieve the targets set for every fiscal.

- Monitoring additions to Archives
- Monitoring RNA Usage
- Web metrics of DNA E-Paper site
- Web metrics of Syndication site
- Syndication Business - Analysis and Revenue Reports
- Brand Reach Analysis

VALUATION OF DIGITAL ARCHIVES AND RNA FUNCTIONS and SERVICES

RNA is essentially a service function and such a function is said to be difficult to quantify. It is qualitative, it is subjective and a matter of perceived value in an organization would be the accepted norm. We run contrary to this. Every RNA service is measurable. The addition of Stories and Images to the DNA Archives every month is monitored. The web metrics of the two global touch points www.3dsyndication.com and <http://epaper.dnaindia.com> are collated every month. The brand reach of DNA on syndication partner sites comes in as Usage Reports from those that we have got licensing agreements with. All of these comprise the monthly MIS which is compiled in to monthly/quarterly/annual reports for the top management. A notional value is attributed to these figures based on a calculation matrix that has been formulized. The sum total of this notional value is the value attributed to RNA.

RNA SERVICES - VALUATION MEASURED IN TERMS OF

- Availability of a comprehensive 24x7 digital database and research facility
- Desktop access across editions
- Improved efficiency and turnaround time of editorial
- Research support for Sales revenue
- A Learning Centre for all employees in the organisation

DIGITAL ARCHIVES - VALUATION MEASURED IN TERMS OF

- Value of Digital Archives being built – E-paper, Stories, Photographs, Cartoons, Illustrations
- DNA Assets valuation

BRAND VALUATION MEASURED IN TERMS OF

- Brand reach generated through Syndication business – second/third party platforms
- Brand reach generated through E-paper and Syndication site

CONCLUSION

DNA – Research N Archives has earned its place in the management mind space by building a successful model of Information and Content Management and Content Monetization. With a small team of 11 members RNA delivers substantial and measurable value to DNA.

DNA – Research N Archives went from Cost centre in 2005 to Profit centre in 2008 and has been posting about 40% year on year (YOY) revenue growth since. On the anvil is a Knowledge Management system. We are partially there but intend to have the full system in place in the very near future

REFERENCES

1. WAN-IFRA 2009: World Press Trends
2. Pujari, Anita and Shaikh Fatimabee. *Content Syndication Business Initiative by a Newspaper Library: a Case Study*. IcoASL 2008. Shaping the Future of Special Libraries – Beyond Boundaries

DIGITAL PRESERVATION AND ACCESS TO PRINT MEDIA RESOURCES: EXPERIENCES AT THE TIMES ARCHIVES AND KNOWLEDGE CENTRE, INDIA

R. Venkata Kesavan

Shankar Marathe

Salila Sreesan

and

Prof. Vijaya P. Rajhansa

ABSTRACT

One of the prime requirements for running a successful newspaper publication is getting access to high-quality information. A journalist is always in need of current as well as archival information and accesses it from a number of sources, and the prime one being an in-house newspaper library or archives. Therefore, collecting, selecting, processing, preserving and providing access to published information is an important and challenging task for the newspaper archives and libraries. This paper deals with how the Times Archives and Knowledge Centre (TAKC) in India, has been preserving various information resources, particularly in digital form and how the access to digital information resources is facilitated for use by the journalists for their reference and information needs. Besides, the paper also highlights some of the practical issues faced by the Times Archives in digitization and preservation. It also attempts to share some of the best practices that are being followed by TAKC in digital preservation and offers a number of suggestions for those who want to take up digitisation projects in newspaper archives and libraries.

Keywords: Newspaper; Libraries; Newspaper Archives; Digital Preservation

INTRODUCTION

Globally, the media sector is always in need of high quality information resources to enhance the knowledge of its employees for creation and dissemination of media content for the purpose of mass consumption. Therefore, apart from preserving the in-house publications, clippings, photographs and other resources for posterity, fulfilling the ever-growing information needs of media industry professionals is a prime responsibility of a media library or archives. Even though Internet has a huge volume of public domain information in the offing, as Williams and Nicholas (1997) point out, journalists just do not have time to worry about the volume of information, authority questions or latest sources of data on the Internet. Besides, multitasking for instance is a trend with increasing acceptance in news media, as journalists have to cope with

a widening range of responsibilities in order to get their jobs done (Saltzis and Dickinson 2008). Therefore, the electronic revolution has brought in dramatic changes in the way newspaper librarians do their jobs and how they are related to the newsrooms (Chepesiuk 1996: 49). Typically in a media organization, the editorial staff (journalists) and many others, who work in revenue generating business departments (such as advertising, marketing and strategic planning etc.), are in need of various kinds of information / data which need to be supplied in time with quality, as timely application of knowledge leads to competitive advantage. This paper profiles the activities, including digitisation of archival resources, of the Times Archives and Knowledge Centre (TAKC), which is one of the oldest service divisions of The Times of India Group, was originally set-up as an 'in-house Reference Department' in 1912.

ABOUT THE TIMES OF INDIA GROUP

Bennett, Coleman & Co. Ltd. popularly known as The Times of India Group is the largest media house in the country that has revolutionized the world of information and entertainment through innovative thinking, winning strategies and global management practices. The 172-year-old Times Group is a market driver across a number of media platforms and it is one of the most respected business houses in India. The Group's leading brands include: newspapers such as *The Times of India*, *The Economic Times*, *Navbharat Times*, *Maharashtra Times*, *Mumbai Mirror*, magazines like *Femina* and *Filmfare* (published in association with BBC). Besides other popular electronic media brands are Radio Mirchi (FM radio network), Zoom - Lifestyle television channel, TIMES NOW - Television News channel, ET NOW-business news channel and Internet portals such as *indiatimes.com*, *TimesJobs.com* etc. The Times of India Group, having a pan India presence, employs over 10,000 people and reaches to millions of consumers through its various media initiatives. In supporting the above business activities, using its vast collection of information resources and expertise, Times Archives and Knowledge Centre (TAKC) plays a very vital role.

TIMES ARCHIVES AND KNOWLEDGE CENTRE

As Mahapatra (2007: 74) notes, it is a universally known fact that the effectiveness of the press largely depends upon the efficient and well organized library services. With this kind of a vision the TAKC, as a newspaper reference dept., was set up over a century ago in Mumbai (then Bombay), and has gradually expanded to Delhi, Ahmedabad, Bangalore, Pune and other cities. The TAKC Mumbai maintains the Central Archival Treasure Trove of the Company with an archival collection of over sixty current and erstwhile publications, the earliest being the flagship publication of the Group, *The Times of India*, published since 3 November 1838. As one of the largest newspaper archives in Asia, TAKC acts as the 'central knowledge centre' of the Group and offers a range of value-added knowledge and information services primarily to the in-house users and often based on special needs caters to external users.

Archival Resources and Collection Management at the TAKC

The TAKC collects, processes and preserves a wide range of media resources in the form of multimedia contents. Some of these resources are dating as far as back as 1830s. *Table 1* provides the broad details of print media resources which are preserved at the Times Archives. The archival collection include The Times of India Group's print publications (preserved in the form of bound volumes, about 5 million pages), microfilm collection of publications (about 1.5 million pages), photographs (over 1 million), clippings of news items and articles tracked from various newspapers and magazines (over 1 million items, covering 6000 subject areas and 11,000 personalities from all walks of life), and a rich collection of original Cartoons, created by eminent cartoonists, particularly of the legendary R.K. Laxman. Collectively all these information resources form as the core knowledge base for the information processing activities within the Group.

Digital Preservation and Archiving

The TAKC is not only a very vital 'resource centre' for editorial teams and other users in its parent organisation, but also for the genuine external users from the country and abroad at large. As Gatos, et al. (2000: 77) opined, it is highly essential for newspaper libraries to digitise its archival resources due to their immense significance for the entire society at large:

"Newspaper archives incorporate a number of facets concerning a country's social, political and economic history. Therefore, they are rightfully considered part of a country's national heritage. Digital preservation of such archives aims both at the salvation of endangered material (paper) as well as the creation of digital library services that will allow full utilization of the archives by all interested parties."

As noted by Venkata Kesavan (2009), a digital library allows a user to preserve, search, retrieve, print, save, share and re-use the archived information instantly. In order to preserve the vital archival resources for posterity as well as faster information storage and retrieval for the ongoing information needs, the TAKC had begun digitisation projects way back in the year 2000 on a small-scale, and since then, as indicated in the *Table 1*, a number of projects have been initiated. About 500 GB of various digitised contents have been preserved in a distributed server system. So far, over 500000 pages from newspapers and magazine collections (including the complete collection of *The Illustrated Weekly of India* – a ceased publication) have been digitised and preserved. For all retrospective publications, their respective microfilm collections have been used as the basic input source for digitisation. However, from mid 1990s, directly exported from the newspaper production systems, the text based, born-digital electronic archives of *The Times of India (1994 onwards)* and *The Economic Times (1995 onwards)* are being maintained in a system called TAS (Times Archival System). This is an in-house developed wide-area network based system, runs on *Oracle*, accessible through individual Log-in IDs, mainly offered to our Editorial Staff across our Offices in India. TAS offers access to over 700000 news items and articles of these two publications (Mumbai & Delhi Editions), covering 1994-2004.

Table: 1 - Collection Status of Print Media Resources Preserved and Accessible at the TAKC

TYPE 1.	PRINT PUBLICATIONS (1838+)
Description	All newspapers and magazines published by the Times Group, such as <i>The Times of India</i> , <i>The Economic Times</i> , <i>Mumbai Mirror</i> , <i>Femina</i> , <i>Filmfare</i> etc. Also includes ceased publication such as <i>The Illustrated Weekly of India</i> (About 60 titles, published since 1838)
Items in Hardcopies	5 Million pages of original issues of various publications in Bound Volumes. (Closed Access through TAKC Staff)
Items in Microfilm	1.5 Million Pages (2 sets of master negatives and one set of reference positives)
Items Digitised	One Million articles and news items published since 1994 till date, in a combination of texts, TIFF images and PDF. Access is facilitated through e-paper archival system, which runs on XML (Extensible Markup Language) based <i>Olive Software</i> . (Free on Internet at: http://epaper.timesofindia.com)
TYPE 2.	PHOTOGRAPHS (1905+)
Description	Collection of photographs, positives and negatives on over 4000 subject areas and 16000 personalities)
Items in Hardcopies	1.2 million items (print photographs, TPs, Positives and negatives) in specific folders shelved in filing cabinets, supported by computerised catalogue to retrieve the folders. (Closed Access only through TAKC Staff)
Items Digitised	200,000 digital images in a 300 DPI JPEG format. Access is facilitated through an online archival system called <i>COMYAN Image System</i> – proprietary content management tool. (access over Intranet through individual Login IDs)
TYPE 3.	BACKGROUND FILES (1900+)*
Description	Published information in the form of clippings tracked from various publications on over 5000 subject areas and 11000 personalities
Items in Hardcopies	17000 folders with over 1 million clippings, classified and arranged based on a customised, in-house developed ‘ <i>alphanumeric classification scheme</i> ’ specific for subject areas and personalities
TYPE 4.	E-CLIPPINGS (1994+)
Description	Clippings of news items and articles tracked on daily basis from over 60 in-house and competitive publications, digitised and indexed. <i>Strictly meant for in-house reference only</i>
Items Digitised	700000 documents (* Also includes over 100, 000 clippings selected from Background files published prior to 1994). Information processing, storage and retrieval are through <i>DataScan Online</i> , a proprietary web-enabled content management tool (Restricted access only through individual Login IDs on Intranet)
TYPE 5.	CARTOONS AND ILLUSTRATIONS (1954+)
Description	Collection of original cartoons and illustration created by leading cartoonists associated with the Group namely Mr. R. K. Laxman, Mr. Mario Miranda and others
Items in Hardcopies	Over 30,000 originals cartoons, caricatures, illustrations and info-graphics that have appeared in various Times Group publications, arranged in specific folders, by cartoonist, publication, subject and date-wise
Items Digitised	20,000 cartoons in TIFF format (150 DPI) with relevant metadata stored in MS-Access. A web enabled information storage and retrieval system development work is in progress. (Restricted Access only through TAKC staff)

As Drennan (1999: 269) observes, the advent of Internet technology heralded the possibility of making information available to as wide an audience as possible; newspapers have exploited these new opportunities to branch into different markets and reach new readers by creating web versions of, or parallel

publications to, their print editions. From 2005, the entire collection of the above said two main publications are available in e-paper archive (<http://epaper.timesofindia.com>) with Internet-based free access for internal users as well as for public. The e-paper archive provides access to complete newspapers (converted from born-digital full issue PDF), as they have been published in hardcopies (with same look and feel – covering news items, articles, images and ads) and is being updated on a daily basis. Our e-paper archive has been developed using *ActivePaper Archive*® of Olive Software (olivesoftware.com), which uses Dublin Core Metadata standard for managing its XML based digital repository. As Jones, et. al. (2006) suggests, mobile and handheld devices are now gradually being used to access e-resources by users, who have fast-paced lifestyles. Sensing such trends, starting year 2007, the Group has launched its m-paper (mobile e-paper), which can be accessed at <http://mobilepaper.timesofindia.com>, by using high-end cellular phones models from BlackBerry, Nokia, Samsung etc.

While these developments were taking place on the one hand, on the other hand post year 2000, the TAKC has also started digitising various precious and high-value photographs, mainly using outsourced resources and partly in-house resources. From year 2004, we are using a proprietary web-based photographs content management tool called *COMYAN Image System* (comyan.com), which facilitates faster image processing (production and printing work-flow process) as well as archival storage and retrieval of photographs including the new ones that are shot on a day-to-day basis, which are born-digital images. In the *COMYAN* photo archive, IPTC (International Press Telecommunications Council) Photo Metadata headers are used as metadata standards, since it is an internationally accepted standard, particularly within the media industry. The digital photo archive currently has over 200,000 digital images (JPEG – 300 DPI) with relevant metadata and is growing by every passing day.

As of now, no one would underestimate the continued use of press clippings in a newspaper library and information services. As Mahapatra (2007: 76) has observed, newspaper cutting collection popularly known as press clippings appears to be at the heart of these news libraries' total collection, and press clippings service assumes greater importance than any other services in the newspaper libraries. While the older press clippings in the TAKC are maintained in hard copy format in subject specific folders, selected clippings (about 700000 documents) published since 1996, on various major subject areas and leading personalities (on the later aspect the coverage is from 1950 onwards in some cases) have been digitised. The digitised images have been saved as single bit B & W BMP format and they are accessible on a proprietary intranet based content management tool called '*DataScan Online*', which uses *MySQL* for the back-end database and *ASP.NET* for front-end applications. This package is used for digitisation as well as archival storage and retrieval of digital contents. Metadata elements in this database are customised according to our local requirements and the key elements are record Id, classification No. clippings type, title, byline, source, date, pagination, subject, and keywords etc. The information resources archived in *DataScan Online* are only meant for internal users for research purposes and strictly adhering to copyright restrictions, no external users are given access to this system.

Since 2006, yet another digitisation project exclusively focused on the Cartoons Collection of TAKC has been started. All the original cartoons, about 20000, created by R.K. Laxman (leading Indian cartoonist, who has been associated with *The Times of India*, since 1954) have been digitised and preserved in our local archives. All relevant metadata pertaining to digitised cartoons are being captured, based on the Dublin Core Metadata Standard. Currently, a full-fledged web enabled, XML based digital archive content management tool is being developed, by benchmarking the features with BCAD (The British Cartoon Archive Digitisation) project, which is one of the best online examples (cartoons.ac.uk) of cartoons digitisation projects.

Archival Preservation

The physical archival storage areas are always kept well ventilated, dust free and pest controlled. As per the internationally recommended archives environmental management standards (eg. BS5454:2000), the required ideal temperature (between 16-19°C) and relative humidity level (between 45-60%) are always maintained in 24x7. As a precautionary measure for digitised data security, a complete back-up of digital resources (in high-end servers) as well as microfilm collection parallel to the main collection of the central archives are being maintained at a remote location, with a similar kind of temperature control and security systems.

Reference Library

TAKC houses a well equipped reference library, which provides access to a special collection of over 10000 books / reports, 100s of current newspapers, magazines and 1000s of CDs, DVDs and other audio visual resources. Besides internal users, on a daily basis a number of external users (eg. students, researchers and govt. depts.) also consult the library by prior appointments.

BUSINESS INFORMATION SERVICES AT TAKC

As newspapers grow in size, as they extend their agencies to encompass more and more of the worlds' interests, and as life becomes more involved, there are new demands placed upon the newspaper reference library, that newspaper reference library which began life as the "morgue" (Desmond 1929: 323). The trend suggested above then is even more relevant in the present context. Apart from supporting editorial activities, the TAKC proactively supports various business divisions (eg. advertising, marketing, business development etc.) of The Times of India Group by offering an exclusive menu of customised value-added information services ranging from that of compiling updates on current developments in the corporate sector to in-depth information analysis of various business sectors, companies and brands. To facilitate business support services, the TAKC has subscribed to a number of useful business resources such as Prowess, C-Line, Prime Database, TAM-Adex, Ebsco etc. As Drennan (1999: 269) suggests, a secure internal network could transform the way in which information is stored, manipulated, distributed and retrieved, meeting the needs not only of the journalistic staff, but also all employees, most of the business information services of TAKC are offered through Timescape (Company's

Intranet) or through secured e-mail IDs of employees. The company website is the primary method of communication between the library and the rest of the corporation, note Rimland and Masuchika (2008). While by default all the internal users have direct access to the TAKC, the external users are permitted only for academic, research, legal and personal uses and permission always granted by prior appointment. For commercial use of TAKC resources, the customers are directed to Times Syndication Service, located at New Delhi.

INFORMATION NEEDS OF JOURNALISTS AND USE OF ARCHIVAL RESOURCES

Any newspaper reference system, of course, must be adapted to meet its own needs (Meyer 1926). At the TAKC, it has been observed that ever since the introduction of Internet in the Organization in late 1990s, the information seeking patterns of journalists have gradually changed, as most of the general reference queries have been solved through the Internet resources by journalists themselves. As authentic information is very vital for the successful journalist, we have been regularly observing and recording the changing information needs of our journalists. Apart from conducting formal group discussions periodically and informal one- to- one reference interviews with the journalists on a day-to-day basis, we also use their e-mail queries as well as feedback messages, as research tools for identifying the information needs and library use patterns. Table 2. presents the summary of the information need queries of journalists solved on specific subject areas. The data has been extracted from the TAKC MIS records for the period 2006-2009.

Table: 2 - Information Needs of Journalists observed at TAKC

SN	Information on major Subject areas of Sought	% of queries
1	Business and Economy	20
2	Political Affairs	15
3	Social activities and culture	15
4	Entertainment & Sports	10
5	Profiles of Personalities	10
6	Govt. Information	9
7	Education and Career	7
8	Religion & spirituality	5
9	Science, technology and trends	5
10	Health	4

Data Source: TAKC – Quarterly MIS Reports from 2006-2009

Use of Archival Resources by Journalists

In their research study on information seeking behaviour of Kuwaiti journalists, Anwar, et al. (2004) found that among the various purposes for which gathered information is used, writing a news item, writing a feature and general knowledge occupied the first three positions among the journalists. This trend is more or less similar among the TAKC users as well. The users (journalists) of the Times Group comprising of reporters, sub-editors, copy editors, feature

writers, news editors, columnists, editors etc. frequently use TAKC resources through visit to the facilities, by sending e-mail requests and getting online access to already digitised resources on the Intranet and other Internet based resources. Table.3 maps the various purposes of use and types of information resources used by journalists at TAKC during 2006-2009.

Table: 3 - Purpose and Use of Archival Information Resources by Journalists at TAKC

SN	Purpose	% of use	Types of resources used
1	Compiling background information	30	Background info. files, clipping database, TAS, e-paper archive, personality profiles, microfilm collection, photographs
2	Preparing special reports	15	Background info. files, clipping database, TAS, e-paper archive, microfilm collection, photographs, current competitive publications
3	Researching for feature articles	10	Magazine articles database, clipping database, TAS, e-paper archive, microfilm collection, photographs, current competitive publications, Internet Resources
5	Researching for news items and follow-up of stories	10	Magazine articles database clipping database, TAS, e-paper archive, current competitive publications, Internet Resources
4	Authentication and verification of factual information	8	Library Reference Resources, Internet Resources, Background files
6	Compiling personality profiles	8	Background info. files, personality profiles, clippings and magazines articles database
7	Referring current affairs and breaking news stories	6	Internet, TV Channels, Current Publications, syndicated news services
8	Writing editorials	5	Background info. files, clipping database, TAS, e-paper archive, microfilm collection, current competitive publications
9	Generating story ideas	5	Current competitive publications, internet resources, e-mail news alerts, syndicated news services
10	Other purposes (self development, journalism professional update, events update etc.)	3	Internet resources, current newspapers and magazines, library books, newsletters, reports etc.

Data Source: TAKC – Quarterly MIS Reports from 2006-2009

BENEFITS AND ADVANTAGES OF DIGITISATION

Digital collections are better because they are dynamic, changeable, constantly updatable and available any time, any place, anywhere, observes Joint (2006). As we gradually move on from maintaining physical archives to digital archives, the new environment offers a number of benefits and advantages to the library staff as well as users. Digitization provides new opportunities to extend their utilization and thus avoid damage to these valuable historical archives (Yu 2007). Modern scanners and digitizers may help not only to preserve material which would otherwise be lost, but to preserve it in machine-readable form, thereby improving access while preserving the text (Sweetland 1992). While our experiences at TAKC also support similar views, we have listed the following

practical benefits and advantages of digitisation, most of which could be common with others as well:

- Preserves all published contents for posterity and future reference
- Saves the time of library staff as well as users in information storage, search and retrieval processes
- Offers simultaneous access to multiple users in 24x7, irrespective of their place and location
- Maximizes the usage of archival resources
- Adds value to editorial research work and new content generation process
- Helps in marketing of library and information services, contents monetization and revenue generation activities
- Saves the library physical storage space
- Overall improves the quality of newspaper library and information services by adding tremendous value to our user groups.

BEST PRACTICES SUGGESTED FOR NEWSPAPER ARCHIVES DIGITISATION PROJECTS

Large and small organisations around the world from many different sectors (museums, archives, libraries, art galleries, government and commercial) have been creating or converting resources into digital form for a wide range of users, notes Holley (2004). As the digitisation is not any more a new concept in the newspaper libraries and archives, as a management philosophy ‘best practices’ that are proven as successful ones from various digitisation projects can be gathered through professional interactions, by reading published literature, attending conferences, seminars, workshops etc. and used in our digitisation projects. One of the workable ideas, as Lanz, et al. (2009) suggest that understanding digitisation requirements is the most important factor in estimating project costs and developing a project plan. It may be noted that, however one tries to equip their library system with best possible technologies and resources, it is always advisable to keep the end users in mind while implementing any digitisation projects. Because, persons who encounter problems while using computers or associated technologies such as internet, e-mail, online access and the like, may resort to avoidance, that is, they may avoid using these technologies and manage without them (Sami and Pangannaiah 2006). In order to avoid any adverse situations like this, libraries at many newspapers have regular training sessions, where journalists from newsroom are taught how to improve their research skills and how to use the Internet and in-house and commercial databases (Chepesiuk 1996: 49). Based on the good amount of successes achieved in a number of digitisation projects at TAKC and learnt from others experiences, we suggest the following best practices to those who want to take up such projects in future:

- Develop a Digital Collection Development Policy and Strategy document, which should be a well thought-out one and it should meet your ‘parent organization’s core mission’. Selection of resources for

digitising should be relevant to your library's present and future user groups

- Selecting an efficient project management team is vital for the success of any digitisation project. Where internal team is not available, it is always better to outsource digitisation work
- While designing a standard metadata schema and information storage retrieval system the library staff and the end-user groups are to be kept in mind, besides matching them with the information resources
- Before getting into a large scale digitisation project 'piloting' is a must, as one learns a lot from such small initiatives
- Using high quality ICT tools always enhances the quality of digital resources and helps in better use, re-use and long-term preservation, therefore, do not compromise on such resources
- Training library staff and educating library users in digital resources management and access respectively will help to reap maximum advantages from such initiatives
- Setting up an effective communication system (formal as well as informal) for interactions within the project team, top management of the organisation and external vendors, if any is crucial for the success of a digitisation project
- From time to time, through scientific methods identifying the changing information needs of the users and continuously evaluating the use of digital resources is a must. A structured monthly MIS reporting of digital resources use will help to maintain up-to-date services as well as justify the implementation of any digitisation projects to the Management
- Finally protection of digital resources using encryption technology, watermark and digital signature etc. can be considered. If, in case of any unforeseen disasters are to be handled, maintaining adequate back-up of digital resources, on site as well as off-site is a must.

EMERGING ISSUES

Even though digital newspaper archives offers a number of benefits, as like in every other LIS projects, the digitisation projects in newspaper archives also bring in a host of problems and issues to be dealt with. One of the critical problems is that computer technology is changing so rapidly that the interfaces, technical standards and file structures which will be used by computers even 20 years into the future are quite likely different from those used today (Cathro 2001: 337). As Davis (2008) has observed, not only is the library as a place changing, but also the nature of information itself is being transformed through the shift from linear text-based information environments to the networked hyperlinked non-linear environment of the Internet. Despite the best efforts from the TAKC for implementing digitisation projects and offering high quality information services by using the archived digital resources, some of our journalists and editorial staff continue to look for information from either their

personal archival collection or some other public domain web-based sources. Because, information users are no longer solely reliant upon physical storehouses of information and the intermediary skills of a librarian to answer enquiries or search for information, notes Tanner (2001). As huge sums are being invested in the creation of electronic content ...attention must be given at all stages of the content creation, storage and delivery process to the digital content formats being employed, and steps must be taken to actively manage content formats over time, to guard against the dangers of creeping technical obsolescence or long-term degradation of resources, cautions Williamson (2005: 512). As Ooghe and Moreels (2009) argue, digital documents serve no purpose unless metadata are linked to them, but creating these is usually time and labour intensive. Besides, such common issues, archival digitisation projects also bring in a number of managerial problems and strategic issues, which one needs to address from time to time. Some of the important ones experienced at TAKC are listed here:

- Budget – Getting adequate funds for digitisation, setting up ICT facilities, management and operational resources are a big challenge for the libraries, even in big media houses like ours
- Metadata schema -Standardizing metadata schema for specific type of print media resources is a never ending issue (eg. Cartoons)
- Manpower -Finding quality manpower and training them in digitisation projects is a continuous issue
- Content Management Tools -There is no single best digital archival system available in the market that would suit all kinds of print resources. For a smaller newspaper library, it is not feasible to develop in- house archival storage and retrieval systems
- Users' Resistance – Some of the journalists' resistance for using digital resources continues to be an issue which calls for some cultural changes at the end-user level
- Legal Issues – Post digitisation of precious archival resources, protecting Copyrights, IPR, fair use of resources etc. are some other challenges one needs to look into
- Technology Issues – In a large scale digitisation projects, there are issues such as finding massive hardware for archival data storage, back-up, developing systemic tools for disaster recovery and maintaining digital archives facilities 24 x 7 etc. also needs to be addressed.

CONCLUSION

However one tries to put in best possible efforts in running a newspaper library, as Whatmore (1994) notes, library's contributions to the success of the media organisation are difficult to evaluate. As every day, the journalists approach the newspaper libraries with a variety of information needs, mostly at the eleventh hour and at times it is really challenging and daunting for the library professionals to keep up with such never ending demands. Perhaps the next generation of librarians and information managers, bring together the skills of

journalism and archivism and add technological skills (Hunt 1998) in order to better equip themselves to remain relevant in the present era. As Venkata Kesavan, et. al. (2008) suggest, one should be willing to explore and experiment new information resources and services in order to meet the ever increasing needs of highly dynamic user groups. Above all, as Martin (1999) reported, our efforts have to be user-based, we need to listen to our users and understand their jobs and need and build a service based around those expectations. In the coming years TAKC will surely continue to engage itself in various digitisation projects in order to play a vital role not only in serving the information needs of the employees of The Times of India Group, which is rapidly changing and diversifying, but also the genuine external users from India and abroad. As Bremer-Laamanen (2006) hopes, during the next few years it is possible for us all, with large newspaper collections on microfilm, to digitise our collections, to build "Digital Newspaper Libraries" for our users. While we are constantly striving to offer high quality information services, to achieve further excellence in our services, at TAKC, we have been involved in a number of on-going digitisation initiatives for preserving the archival treasure trove of The Times of India Group, such as digitisation of old photographs, cartoons, illustrations etc. As King (2005) states, it is possible now to plan to have microfilms scanned digitally, with the digital images capable of further manipulation by software, to permit a good degree of readability, with the texts also being searchable. However, as Sims (2008) warns, it is to be noted that newspaper digitisation projects involve painstaking preparation and would call for labour intensive work to accommodate the varied sizes of the same newspaper, as well as fragile conditions. Nevertheless based on the already achieved successes and experiences gained in small-scale digitising projects, in December 2009, we launched a mega project of digitising the entire collection of *The Times of India*, published since 1838 till 2004. About one million pages are to be digitised and integrated with the existing e-paper archive. As the entire digitisation process is completely outsourced, we hope to complete the project by end 2010. Surely such initiatives hold a lot of professional challenges and opportunities for the future.

REFERENCES

1. Anwar, M. A., Al-Ansari, H. and Abdullah, A. 2004
Information seeking behaviour of Kuwaiti journalists
Libri, 54(4): 228–236
2. Bremer-Laamanen, M, 2006
A Nordic digital newspaper library
In *International Newspaper Librarianship for the 21st Century*, pp. 251–256, edited by Walravens, Hartmut (IFLA Publications), Berlin, New York (Walter de Gruyter – K. G. Saur). 298p.
3. Cathro, W. S. 2001
Keynote paper: visions for fundamental change in libraries and librarianship for Asia Pacific
Library Review, 50(7/8): 334 – 342.
4. Chepesiuk, R. 1996
The electronic news library: an online revolution takes shape

- American Libraries*, 27(9): 48-50.
5. Davis, C. 2008
Librarianship in the 21st century-crisis or transformation?
Public Library Quarterly, 27(1): 57 – 82.
 6. Desmond, R. W. 1929
Instruction in newspaper library methods
Special Libraries. 20(9): 323-325.
 7. Drennan, J. 1999
The introduction of intranets into the newspaper industry
Aslib Proceedings, 51(8): 269-274.
 8. Gatos, B. et al. 2000
Automatic page analysis for the creation of a digital library from newspaper archives
International Journal on Digital Libraries, 3(1): 77–84
 9. Holley, R. 2004
Developing a digitisation framework for your organisation
The Electronic Library, 22(6): 518-522
 10. Hunt, R. 1998
News, information and value.
Aslib proceedings : New information perspectives, 50(8): 215-220.
 11. Joint, N. 2006
Digital library futures: collection development or collection preservation?
Library Review, 55(5): 285-290.
 12. Jones, M., et. al. 2006
Changing the pace of search: supporting “background” information seeking
Journal of the American Society for Information Science and Technology, 57(6):838–842.
 13. King, E. 2005
Digitisation of newspapers at the British Library
The Serials Librarian, 49(1): 165 – 181.
 14. Lanz, D. et al. 2009
The new Papers Past: an international collaboration between New Zealand, India, Germany, and the United States
OCLC Systems & Services: International Digital Library Perspectives, 25(4): 287-294.
 15. Mahapatra, R. K. 2007
Newspaper libraries in Orissa
Orissa Review, February-March: 74-77.
 16. Martin, H. 1999
The changing information environment in the media: case study The Guardian/Observer
Aslib Proceeding, 51(3): 91-96.
 17. Meyer, R. 1926
Efficiency in the newspaper library
Special Libraries, 17(9): 369
 18. Ooghe, B. and Moreels, D. 2009

- Analysing selection for digitisation: current practices and common incentives
D-Lib Magazine, 15(9/10).
<http://www.dlib.org/dlib/september09/ooghe/09ooghe.html> (Accessed on 15 January 2010)
19. Rimland, E. and Masuchika, G. 2008
Transitioning to corporate librarianship
Journal of Business & Finance Librarianship, 13(3): 321 - 334.
 20. Sami, L. K. and Pangannaiah, N. B. 2006
Technostress: a literature survey on the effect of information technology on library users
Library Review, 55(7): 429-439
 21. Saltzis, K and Dickinson, R. 2008
Inside the changing newsroom: journalists' responses to media convergence
Aslib Proceedings: New Information Perspectives, 60(3):216-228.
 22. Sims, J. 2008
Librarianship in the 21st Century: a British Library perspective
Legal Information Management, 8(2): 84-91
 23. Sweetland, J. H. 1992
Humanists, libraries, electronic publishing, and the future
Library Trends, 40(4): 781-803
 24. Tanner, S. 2001
Librarians in the digital age: planning digitisation projects
Program, 35(4): 327-337
 25. Venkata Kesavan, R., Sreesan, S. and Silveira, S. 2008
Business information and knowledge support services at the Times Archives and Knowledge Centre: a case study, pp. 487 - 497.
In *Proceedings of the 9th Annual National Convention of MANLIBNET on Business and Management Librarianship: The Decade Ahead*. ICSSR, New Delhi, February, 4 - 6, 2008. 540p.
 26. Venkata Kesavan, R. 2009
Digital library services: a practical approach for collection development, organization and management
Journal of Lib. Inf. & Comm. Technology, 1(1): 5-24.
 27. Whatmore, G. 1994
Getting the editor's ear: the Manchester Guardian Library in the 1950s
Library Review, 43(3):8-16.
 28. Williams, P. and Nicholas, D. 1997
Journalists, news librarians and the Internet
New Library World, 98(1137): 217-223.
 29. Williamson, A. 2005
Digital directions: strategies for managing digital content formats
Library Review, 54(9): 508-513.
 30. Yu, S. 2007
Constructing a DRM framework for historical newspaper archives
The Electronic Library, 25(6): 778-789.

NEWSPAPER DATABASE MANAGEMENT AT MICA'S KNOWLEDGE EXCHANGE & INFORMATION CENTRE

Dr. Shailesh R. Yagnik

Niraj R. Patel

and

Lavji N. Zala

ABSTRACT

In absence of relevant and consistent journals and magazines related to advertising, mass media and marketing, the newspapers including business and general interest newspapers become vital source of information for advertising and related areas. This paper focuses the history and development of newspapers in India. Knowledge Exchange & Information Centre (KEIC) has designed the Newspaper Clippings Database using WINISIS and GENISIS by creating many fields for the retrieval of information. Database management at KEIC has been briefly mentioned. This paper also gives information on copyright and how the database should be maintained within the terms and conditions of different newspapers. It also provides year wise input of data, newspaper wise break-up, and content analysis of 1,58,239 records and highlights of a users survey. How many newspapers provide free access to their news items and print ads in the form of e-paper and archive has also been discussed in the paper. Lastly it discusses the manpower, ICT and future of newspaper clippings.

Keywords: Newspaper Database, Copyright, Users Satisfaction, Content Analysis, Open or free access, Future

HISTORY AND DEVELOPMENT OF NEWSPAPERS IN INDIA

Johann Gutenberg invented the printing form of movable type in 1440 in Germany. Printing was the solid foundation for information, education, and advertising. Printing was the foundation for increasing literacy and advertising. Until the printing was not started, reading and writing was confined to a certain class. The biggest advantage of printing was making the knowledge available far and wide and within reach of many.

Portuguese brought the first press in 1556 to India but it was used for printing the religious literature and not for advertising. James Hicky started the first newspaper The Bengal Gazette on January 29, 1780 from Calcutta. It is very difficult to think about the newspaper without advertisements. The first issue carried few informative advertisements. Then came *Bengal Journal* published in 1785 and it offered free advertising to Government. After six years in 1786 there were four weekly newspapers published in Calcutta. During the 18th century, advertising in newspapers and magazines was of informative

announcements like births, deaths, arrivals of ships from England, sale of household furniture, etc. *The Courier* was started in 1790 from Bombay and unlike the modern newspapers; it carried advertisements in Marathi, Gujarati, Konkani, Urdu, and Kannada. All the government advertisements were printed in the gazette that was started in 1791.

By 1830 around three dozen English dailies and periodicals were being published on regular basis. The first Indian language newspaper was published in 1833. By mid-century the number rose to a hundred, with ten from North India, seventeen from Bombay and rest seventy-three from the south. In the beginning, neither English nor vernacular press got sufficient advertisement revenue. Due to industrial revolution, mass production was happening and companies were looking for the buyers, therefore, advertising from British business houses rose sharply.

India became independent on 15th August 1947. Government put the ban on import of several consumer goods and encouraged local manufacturers. Competitions among the local manufacturers stimulated the advertising, marketing research and consumer research. Commercial art was very heavily influenced by the West because of dominance of foreign agencies in India. However, slowly Indian agencies were coming out of the western influence and started showing the Indian creative ability.

During post-print era and after the independence, the real growth of advertising in print medium started after 1974 onwards. Before 1974 the advertising expenditure was very low because of low competition in the Indian market. Print medium covers newspapers and periodicals. The frequency could be daily, tri and bi weekly, monthly, bi-monthly, quarterly, semi annually and annual publications. Total 78,223 publications are registered with Registrar of Newspaper of India (2007) out of which 6,826 are dailies and 71,307 are magazines. Advertising industry is always more interested in circulation and readerships rather than number of registered newspapers and magazines in the country. It uses the Audit Bureau of Circulation, National Readership Survey (NRS) and Indian Readership Survey (NRS) data for media planning.

Table 1: Top Five Newspapers

Daily (Eng.)	Daily (Non-eng.)
The Times of India	Dainik Jagran
The Hindustan Times	Dainik Bhaskar
The Hindu	Malayala Manorama
Deccan Herald	Hindustan
The Telegraph	Amar Ujala

Source: IRS 2007 R2

Print medium has its strengths and weaknesses. Advertisers and readers are having limited choice like international newspapers, national newspapers, and local newspapers. Selection of media is taking place based on brand meant for which customer, readership, circulation, frequency, etc.

Press – Strengths

1. Message received at home in a relaxed atmosphere
2. Newspapers have a better shelf life and can be read at leisure and referred to when require
3. The urge to seek news puts the newspaper on a better position to be trusted
4. Reading a newspaper becomes a matter of habit, hence it ensures regular attention
5. Written word had more credibility
6. Message carry urgency
7. High on national coverage

Press – Weaknesses

1. Suffers from the literary barrier
2. Newspaper become stale in a day
3. Lacks drama and emotions
4. Demonstration of product features not effective
5. Overtaken by television in speed
6. Lacks empathetic readership
7. 'Bad' news often considered good news
8. Average time devoted to newspaper reading is very low, hence advertisements do not stand much change of being noticed

Local/Regional Newspapers - Strengths

1. Strong reader loyalty
2. Local coverage
3. Regional flexibility

Local/Regional Newspapers – Weaknesses

1. Small circulation
2. Potential to be biased
3. Generally poor readership data

DATABASE AT MICA-KEIC

Database management for newspapers is a challenging job. It requires skills, consistency, committed staff, right selection of articles/news items on actual and potential value of information. The addition and deletion of subjects depends on the requirements of the institution. Mudra Institute of Communications, Ahmedabad (MICA) is a premier institute in the field of communications management. Knowledge Exchange & Information Centre (KEIC) of MICA has created a vast and rich database on Advertising, Mass Media, and Marketing in the country. Another database from newspapers is on print ads for users. This

database has 6K records. These two databases have filled up the biggest void in the realm of historical data on advertising. The newspaper database has been collated from April 1994 onwards and it contains 157K records as on November 19, 2009.

NEWSPAPERS AS INFORMATION SOURCE

Newspaper has been used as one of the print mediums to release current information to the society on various aspects such as politics, economics, policies, science, health, business, sports, environment, crimes, people, culture etc. A newspaper article has less research value compared to journal article, but in absence of information not available in magazines and journals the content value of newspaper goes up. The newspaper is an important medium so lots of researches are conducted on content analysis and readership of newspapers.

Information on advertising, mass, media and marketing in India are dominated by business newspapers such as The Economic Times, Financial Express, Business Standard, The Hindu Business Line, and Mint compared to magazines. Advertising & Marketing (A&M), Brief, USP Age have stopped publishing and current magazines are Brand Reporter, Impact, Campaign India, Advertising Express, Pitch etc.

The Economic Times, Financial Express, Business Standard, and Business Line have supplements called Brand Equity, Advertising and Marketing, The Strategist, and Brandline (formerly Catalyst) on advertising and marketing respectively. The Pioneer has a page on Unique Selling Propositions (USP). Other papers carry the information in their business columns. The Telegraph used to publish a page on advertising once in a week but later on it was discontinued. The Financial Express has introduced Brand Wagon and has withdrawn it twice, but it was giving very good coverage on advertising.

COPYRIGHT

The permission was sought from all the newspapers press for internal use of clippings and if agencies request, the information will be given to them at a reasonable service cost. The Pioneer, Indian Express Newspapers (Bombay), Ltd., The Hindu, Business Standard, The Economic Times, The Hindustan Times, and The Business Line gave the permission to 1994. The advent of new media after 1992 has impacted the Press to come with e-papers along with their editions. The electronic copyright says that saved information on disk or to any other storage medium should be used for only personal use. Database creation is not permissible by them without prior permission. More the e- newspapers are from the Press more the control of copyright from them on users.

DATABASE FIELDS CREATION

Database creation requires careful selection of the fields such as 1) Author, 2) Title, 3) Newspaper's Name, 4) Date, 5) Subject, 6) Product, 7) Company's Name, 8) Brand Name, 9) Keywords, 10) Agency Name, 11) Text-1 and Text-2.

The fields one to ten were created to access the information from users point of views. Text field was added to the database when most of the Press introduced the archive, e-papers and allowed aggregators to put their full text on the databases. The fields from five to ten become inactive because of full-text search in the database.

For maintaining consistency in retrieving the information the indexer was following guidelines developed by KEIC such as tie-up, joint venture, merger, takeover, expansion, diversification, account gains (e.g. when an ad agency got the account for Lux), product launch, product re-launch, pre-launch, product performance (e.g. Rasna), company performance (Pioma Ind.), market (soft drink), marketing (theory & practice), advertising (theory & practice), mass media (theory & practice) and people profile.

The database was launched on April 1st 1994. It started with 1695 clippings per annum and reached to maximum 20573 in 2007. During 2007, 2008, and 2009 number records crossed 2000 clippings per annum. During 2000 to 2002 the KEIC bought another complete newspapers and magazines full-text database so during that period the number of newspaper clippings had gone down for few papers.

Table 2: Year wise Newspaper Clippings

Year	Number of Clippings	Cumulative Clippings
1994	1695	1695
1995	4212	5907
1996	4430	10337
1997	6172	16509
1998	9579	26088
1999	9056	35144
2000	4643	38787
2001	4262	44049
2002	4535	48584
2003	8124	56708
2004	13684	70356
2005	12224	82580
2006	16533	99113
2007	20573	119686
2008	20208	139894
2009	18345	158239

Between 1994 and 2009 the increase in number of clippings from 1695 to 18345, this shows rise in 9.82 times more compared to 1994. The different subjects have increased due to new development. Initially the print medium was the only active medium but today we have Out of home (OOH), radio, television, online, gaming, mobile etc. Literature related to market, market players, social impact, consumer, legal issues, licences, advertising, marketing, product launches etc. – all these have led to information explosion. The new subscription of existing newspapers and new launches of newspapers such as Mint, and DNA, Money have also added to the numbers. A growing economy

has also impacted the market of foreign companies, and foreign brands. Same way Indian companies and Indian brands are turning into MNCs such as Amul, Tata, Infosys, Mahindra and Mahindra etc. The world has become a global village so users need to access more international information. Newspapers like The Financial Times and The Wall Street Journal are read for information on the world economy and markets.

Table 3: Newspaper Wise break-up:

Newspaper	Starting Date	Closing Date	Records
The Financial Express	19940403		32296
Business Line	19940407		30821
The Economic Times	19949496		28247
Business Standard	19940403		18131
Daily News Analysis (DNA)	20051101		11753
The Times of India	19940413		10329
The Hindustan Times	19940403		5638
The Telegraph	19940814		3929
Mint	20070723		3469
The Pioneer	19970306	20090214	3291
The Indian Express	19940404		3093
The Hindu	19940503		2897
The Observer of B&P	19980108	20001130	1927
The Asian Age	19970202	20051229	1267
The Financial Times	20070816		562
The Wall Street Journal	20051109		546
Others	---	---	39
The Statesman	19950506	19990530	4
		Total Records	158239

The break-up of 1,58,239 newspaper records out of which 1,41,908 clippings have appeared in business papers such as The Financial Express, Business Line, The Economic Times, Business Standard, DNA Money, Mint, and The Observer of B&P. Foreign business papers The Financial Times and The Wall Street Journal have coverage of 1108 clippings on India only. General newspapers have 15,223 clippings which is less than 10% of the total clippings. In others category dates are not given because of high variations and different sources.

This database is a vital source of information on brands, channels, product categories, consumers, market share, market size, advertisers, advertising agencies, media houses, MR agencies. These two databases have supported students, faculty members, and industry.

CONTENT ANALYSIS

Content analysis of 1,58,239 news titles was done with the help of WINISIS, Notepad and SAS. The software has analysed 9,32,270 words, symbols and numbers. From these words, the unique words frequency analysed number 55,077 words. A word appeared more than 500 times was only considered for

the frequency list. There are 134 words appeared more than 500 times. These words are related to geography, marketing, media, consumer, and advertising. For example advertising is dealing with brand, so some reputed brands like TATA, Star, Reliance, HUL, Godrej, Zee, Sony, Coke, Pepsi, Bharati, LG, and Samsung have appeared in the list. Another example is related to communication, it covers TV, media, mobile, channels, online resources, radio, telecom, PC, digital, music, DTH, cable, television, networks, Internet, and Web. Marketing related words are brand, market, retail, ad, launches, sales, launch, brands, business, consumer, products, campaign, price, deal, buy, prices, service, share, services, offer, FMCG, strategy, retailers, segment, cut, rise, women, consumers, and kids.

Table 4: Frequency of words appeared in the newspaper database

Name	Freq.	Name	Freq.	Name	Freq.	Name	Freq.	Name	Freq.	Name	Freq.
India	6734	Products	1087	Buy	848	Offer	676	Ties	564	Push	517
Brand	3889	Eyes	1078	Govt/ Govt.	846	Future	676	Foray	563	Crore	517
Tv	3478	Tata	1051	Small	827	Pepsi	674	Cut	562	Still	516
Market	3339	Range	1048	Prices	821	Fmcg	672	Rise	558	Care	507
Plans	3336	Unveils	1034	Sector	812	Bags	669	Enter	557	Talks	506
Retail	3285	Firms	1007	Way	806	Right	667	Tea	555	Times	505
Ad	2903	Star	1003	Biz	803	News	666	Companie s	554	Better	504
Indian	2706	Reliance	998	Year	800	Strategy	661	Television	549	Invest	503
Launches	2558	Hil/hul	990	Good	796	Retailers	659	Best	549	Account	502
Sales	2554	Car	972	Service	786	Next	659	Women	543		
Growth	2111	Group	971	Study	755	Digital	654	Consumer s	543		
Launch	1935	Home	927	Share	750	Sell	652	Auto	543		
Media	1849	Channels	920	Radio	750	Stake	648	Network	540		
Brands	1847	Health	919	Godrej	738	Down	640	Markets	539		
Business	1751	Online	913	Life	733	Stores	633	Rolls	535		
Global	1630	Dd/doorda rshan	909	Teleco m	730	Open	632	Internet	535		
Mobile	1599	Campaign	909	Pc	715	Music	627	Lg	534		
Marketing	1510	Plan	904	Zee	711	Dth	623	Samsung	533		
Ads	1439	Food	892	Coke	707	China	611	Kids	527		
Us	1431	Channel	890	Sony	703	Cable	608	Change	527		
Industry	1397	Focus	872	Air	697	Educatio n	601	Show	525		
World	1282	Price	865	Indias	695	Drive	592	Web	523		
Consumer	1153	Deal	855	Service s	685	Expansi on	587	Pay	523		
Advertising	1147	Power	853	Hit	680	Bharti	579	Foreign	518		
Rural	1138	High	850	War	677	Segment	568	Expand	518		

The word India appears first in the list. The word India has appeared 6734 times in news titles. Second word brand has 3889 frequency which is 57% of the word India. The first two have major frequency gap otherwise gap is normal with other words.

USERS SURVEY

KEIC conducts the users' survey of its resources for print and electronic databases. This feedback is from outgoing students from all the programmes. The feedback shows that satisfaction of the users on newspaper clippings

database has gone down thrice only in last eleven years below 70 percent. To maintain the newspaper clippings database it requires huge time and commitment. Sometimes most of the information is available in other databases such as ISI Emerging Markets but students prefer to use this database because it gives all the details at one place from different papers and subjects. .

Table 5: Usage of Newspaper Clippings Database and Print Ads by Various Programmes in Percentage

Year	PGPCM (Newspaper clippings)	RCM (Newspaper clippings)	PGPCMX (Newspaper clippings)	PGPCM (Print Ad Guard book)	PGPCM (Print Ad database)
1997-98	89	-	-	89	NA
1998-99	98	-	-	64	NA
1999-00	97	-	-	79	26
2000-01	90	-	-	39	88
2001-02	91	-	-	60	86
2002-03	83	-	-	63	91
2003-04	69	79	-	57	65
2004-05	78	-	-	59	70
2005-06	67	-	-	30	51
2007-08	78	89	60	47	78
2008-09	87	-	-	49	84

PGPCM: Postgraduate Programme in Communications Management

Along with newspaper, print advertisements are also important to users. In the beginning KEIC maintained guard book physically but with the help of e-paper and scanner, KEIC maintains print ad database. The users are more satisfied with print ad database because it is searchable and digital copies are available which has more clarity. Table 5 shows that newspaper clippings database has more usage compared to Print Ad Guard Book and Print Ad Database.

Crafting Creative Communications (CCC) programme is a six months programme. The separate table is given here on users' satisfaction because of two programmes conducted in a year.

Table 6:- Users Satisfaction of CCC Students in Percentage:

Year	Print Ad Guard Book	Print Ad Database
2001	90	85
2002	89	93
2002-03	91	73
2003	100	100
2003-04	100	100
2004	80	80
2005	87	83
2006	NA	NA
2007	NA	NA
2008	63	79

Crafting creative course students are learning copywriting and creative aspects of advertising. They need more information on print ads. Both the forms print and electronic are used by the users equally. Maximum variation in usage could be seen in the year 2002-03

ACCESS OF E-PAPERS AND ARCHIVE OF NEWSPAPERS IN CURRENT ENVIRONMENT

Previously newspapers were available only in print edition. But the current scenario has been changed. Now most of newspapers are accessible through E-paper or archive or mobile edition on Kindle through concerned websites of the newspapers. E-paper allows users to view or access the entire newspaper in its original (newspaper view) format. Whatever content published in print edition such as news items, images, print ads etc are accessible through E-paper. The archive contains selected or entire news items appeared in newspaper. Users can access the news item in textual format but not in Newspaper View format. Most of newspapers' current content (current date) are accessible free of cost. But for the e-paper and archive access, different condition and subscription are applied by newspapers. For example, The Times of India and The Economic Times (all editions) are accessible (e-paper) free of cost from December 2003 and September 2004 respectively. The e-paper of The Business Line and The Hindu are accessible through subscription (See Annexure). Every newspaper has its own terms and conditions in providing access to its digital contents or archives. Librarian should work within the terms and conditions of different newspapers and respect the Copyright Act.

MANPOWER INVOLVEMENT AND FUTURE

Information and Communication Technology (ICT) has reduced the burden of maintaining clippings because previously all the information was maintained manually by press and users. Newspaper database has three stages. First one is physical maintenance of newspaper clippings, either in paper files, box files, and covert clippings into A4 size papers, photocopy the same and convert it into a book form. Second stage is digitization with the help of scanner and third stage is digital copy available on the website of the newspapers but an institute has to maintain relevant information related to areas of study.

CONCLUSION

Newspaper clippings requirements are dispersed and scatted in various newspapers. The librarian has to take the decision regarding cost-benefit analysis between e-papers, archives, commercial databases, and library's own newspaper clippings system. Every organisation has to maintain own clippings for their related areas and on them. Average newspaper life is not more than ten years so accordingly how to preserve the print as well as digital information with least manpower cost and best time management is an important consideration.

Annexure: Free and subscription based access of newspapers

Name of Newspaper	Print Form	e-paper	Open access of print ads	Open Access of E-paper	Open Access of Archive	Subscription is required YES/NO
Asian Wall Street Journal	YES	YES	NO	NO	NO	YES E-paper can be accessed
Business Line	YES	YES	NO	NO	YES From Jan 2000 to Current- Selected news items only	YES E-paper can be accessed
Business Standard	YES	YES	YES Free Registration based access for last 60 days	YES Free Registration based access for 365 days (Request to be sent for the e-paper older than 60 days)	NO Subscription is required to access it (stories older than 90 days and dating back to 1997 are accessible)	YES Subscription is required for e-paper and archive
DNA	YES	YES	YES Only for 365 days Registration is required	YES Only for 365 days Registration is required	YES Only for 365 days Registration is required	YES Older than 365 days e-paper and archive to be subscribed
The Economic Times	YES	YES	YES Free registration is required. E-paper is accessible from September 2004 onwards	YES Free registration is required. E-paper is accessible from September 2004 onwards	YES Free registration is required. Archive is accessible from September 2004 onwards	NO subscription is not required
Financial Express	YES	YES	Registration is required. E-paper is accessible for current 10 days	Registration is required. E-paper is accessible for current 10 days	YES From 5 th June 2007 to till date	Free access for last 15 days only (Subscription charge is not mentioned)
Financial Times	YES	YES	NO	NO	NO	YES Print subscription of FT is mandatory
The Hindu	YES	YES	NO (Images are accessible for viewing purpose but it is price for commercial usage)	NO	YES 01 Jan 2000 to till date	YES Subscription is required to access E-paper
The Hindustan Times	YES	YES	E-paper is accessible for last 7 days.	E-paper is accessible for last 7 days.	NO	YES Subscription is required to access archives from

						October 2004 onwards
The Indian Express	YES	YES	Registration is required. E-paper is accessible for current 10 days	Registration is required. E-paper is accessible for current 10 days	YES Free access from 01 May 1997 to current	Free access of e-paper for last 15 days only. (Subscription charge is not mentioned)
The Pioneer	YES	YES	E-paper is accessible for current 16 days. Registration is required.	E-paper is accessible for current 10 days. Registration is required.	YES Free access of selected articles from January 2006 to till date	NO Free access for last 10 days for e-paper and for archive from Jan 2006 to till date
The Telegraph	YES	YES	Free registration is required. E-paper is accessible from September 2008 onwards			NO Free registration is required. E-paper is accessible from September 2008 onwards
The Times of India	YES	YES	Registration is required. E-paper is accessible from December 2003 onwards	Registration is required. E-paper is accessible from December 2003 onwards	Registration is required. Archive is accessible from December 2003 onwards	NO Subscription is not required
MINT	YES	YES	Free registration is required. E-paper is accessible from January 2007 to till date (limited months and limited days within the months)	Free registration is required. E-paper is accessible from January 2007 to till date (limited months and limited days within the months)	NO Archive section is not available at the website	Free registration is required. E-paper is accessible from January 2007 to till date (limited months and limited days within the months) Subscription details are not given.

REFERENCE

1. Asian Wall Street Journal. <http://www.wsj-asia.com/> (accessed on 02 February 2010)
2. The Business Line. <http://www.thehindubusinessline.com/> (accessed on 02 February 2010)
3. Business Standard. <http://www.business-standard.com/india/> (accessed on 02 February 2010)
4. DNA. <http://www.dnaindia.com/> (accessed on 02 February 2010)

5. The Economic Times. <http://economictimes.indiatimes.com/> (accessed on 03 February 2010)
6. The Financial Express. <http://www.financialexpress.com/> (accessed on 03 February 2010)
7. Financial Times. <http://www.ft.com/home/asia> (accessed on 03 February 2010)
8. The Hindu. <http://www.hinduonnet.com/> (accessed on 03 February 2010)
9. Hindustan Times. <http://www.hindustantimes.com/> (accessed on 03 February 2010)
10. The Indian Express. <http://www.indianexpress.com/> (accessed on 03 February 2010)
11. Mint. <http://www.livemint.com/Lounge.aspx> (accessed on 03 February 2010)
12. The Pioneer. <http://www.dailypioneer.com/> (accessed on 03 February 2010)
13. The Telegraph. <http://www.telegraphindia.com/section/frontpage/index.jsp> (accessed on 03 February 2010)
14. The Times of India. <http://timesofindia.indiatimes.com/> (accessed on 03 February 2010)

ONLINE NEWSPAPER READING HABITS AMONG PhD STUDENTS AND FACULTY MEMBERS IN ALIGARH MUSLIM UNIVERSITY

Dr. Naushad Ali P.M

and

Mohamed Musthafa K

ABSTRACT

The paper defines an online/web newspaper and discusses the features and limitations of this new digital resource. The study reveals majority of the students and faculty members of Aligarh Muslim University are online readers and many of them prefer their regional languages - Hindi and Urdu. Respondents from other than Northern India prefer their own respective mother tongue. The problem of regional language character display has been mitigated to some extent in the last few years. Features like video clips and audio attachments are interesting features of online newspapers. Instant news updates makes online newspapers more attractive to the readers. The favourite items in online newspaper are the local news and comments by the readers.

The data for this study was collected through a questionnaire distributed among the university faculty members and PhD students. An empirical research design along with random sampling technique was used to select the sample.

Key words: online newspaper, electronic newspaper, reading habit.

INTRODUCTION

An online newspaper, also known as a web newspaper, is a newspaper that exists on the World Wide Web or Internet, either separately or as an online version of a printed periodical. Online newspapers are much like hard-copy newspapers and have the same legal boundaries, such as laws regarding libel, privacy and copyright (Wiki 2010). According to American Journalism Review NewsLink, the number of online newspapers rose from very few to 1,300 in 1996 and it was reported that there were 3622 in 1997 world wide. The Internet version of the traditional media of the Fourth Estate attracts thousands of new readers day by day around the world. Unique facilities provided by the online newspapers such as navigability, hyperlink and multimedia entice the readers to this media of daily light reading. A survey conducted by Newspaper Association of America (NAA) points to a tremendous growth in number of online newspaper readers. Newspaper web sites attracted more than 73.3 million monthly unique visitors on an average (43.6 percent of all internet users) in the first quarter of 2009, a record number that reflects a 10.5 percent increase over the same period a year ago (NAA report, 2009). Newspapers, as key vehicles for the development of community in modern society, have been extraordinarily

interested in the notion of building virtual communities as they move their operations online.

Features of Online Newspapers (ONP)

The salient features of online newspapers are:

- Multimedia is one of the unique features of online newspapers. It provides live news/ information along with link to other media like sound records and video clips.
- Highlighted key words with hyperlinks leads the readers to related news items at a single click
- Readers can contact with editors and reporters via e-mail and respond to the columns and news items through Forums and Chat rooms.
- Online newspapers provide easy access to archival collections of past issues which is difficult and tiring in print newspapers.
- Online newspapers update information 24 hours in a day so that instant news coverage is possible.
- Full screen photos and interactive polls that users can participate in are unique facilities provided by online newspapers.

Limitations of Online Newspapers

- Online newspapers need power connection to use an electronic device such as computer to read news; the unstable power connection may cause the failure of the system and finally keep the users aloof from the online newspaper.
- It needs a live internet connection. The feasibility of online newspaper reading is depending upon the speed of internet connectivity.
- An online newspaper is not easily portable compared to print counterpart.
- The style of presentation is different in each online newspaper which may not satisfy the majority of readers and the style itself is being changed frequently.
- Readers need to click and scroll to select a news or article and should be familiar with the search techniques.
- Too many links cause confusion and mislead the reader from his/her interest of reading and waste time.

Objectives of the study

The present study aims to find out the online newspaper reading habits of faculty members and PhD students of Aligarh Muslim University (AMU), India. The areas discussed in this study are the frequency and time spent for reading online newspapers, preferred languages, reasons for choosing online newspapers, media, and locations where from the respondents read the same. The investigators also had objectives to find out the areas of news items which are preferred by the readers and to examine the special features like video clips, sound records, online surveys, hyperlinks, forums and chat room facilities, and

readers' attitude towards advertisements published in online newspapers. Other objectives were to find out the difficulties in reading online newspapers and readers' suggestions to improve the online reading habit.

Scope and limitation of the study

The present study was conducted among the faculty members and PhD students of Aligarh Muslim University (AMU) only. Large groups of AMU community other than these two groups, who are more in number, are not included. The study was focused on online newspapers only. Online editions of TV channels, Radio, and Websites, which are more popular digital mass media (Nielsen online survey, March 2009), were also excluded. Many users do not want to distinguish between newspapers and newssites. According to Nielsen survey sites such as Google News (13,508,000 visitors), Yahoo! News (37,902,000 visitors) and online version of news channels like CNN Digital Network (38,724,000) and Fox News Digital Network (16,791,000) are more popular among the readers of online news.

Need and significance of the study

At the beginning of 1996, there were 154 online papers (Meyer 1996). As of October 25, 1996, the *Editor and Publisher Interactive Online Newspaper Database* had 1562 online newspaper entries. The number of online newspapers on the Web in October of 1996 was 1441. Meyer (1998) reported there were 3,622 online newspapers midyear in 1997 and that number was projected to increase to 4,000 by the end of that year.

The number of newspapers and readers are growing not only in developed countries but also in developing and third world countries. Due to the exponential growth of internet and information and communication technology (ICT) awareness and training inspire the readers, especially the young generation to move into the online version of newspapers. As a place of higher education with Wi-Fi technology, Aligarh Muslim University (AMU) is one of the premier institutes in India to adopt the latest trends and technology for teaching-learning processes. A study about the newspaper reading among the AMU community is important in the era of digital information.

Methodology

The study on online newspaper reading habit was carried out by using a standardized questionnaire administered among the faculty members and bonafide PhD students in AMU. Two hundred online questionnaires were sent by e-mail to the personal e-mail IDs of faculty members and PhD students through online forums and social networking sites used by the student community in the campus. Simultaneously, a printed questionnaire was also distributed among faculty members and PhD students.

Ninety filled up questionnaires were received from faculty members. Out of this, 10 were not considerable due to insufficient data. Similarly out of 95 questionnaires collected from PhD students, 88 were considered for analysis.

In addition to the questionnaire a semi-structured oral interview was conducted for those faculty members and PhD students who did not fill up the questionnaire to supplement the outcome of the formal questionnaire survey. A personal interview with the in-charge of computer section in Maulana Azad Library (university central library) and central computer centre also was conducted. The data were collected during the December 2009 and January 2010.

Analysis of data and Interpretation

Table 1: Online Newspaper reading habit

Category	Category of Respondents	
	Faculty Members	PhD Students
Who read Online Newspapers	71 (88.75)	84 (95.45)
Who do not read Online Newspapers	9 (11.25)	4 (4.54)
Total	80 (100)	88 (100)

(Figures in parentheses denotes percentage)

Table 2: Preference of newspapers

Category of Respondents	Category of Newspaper	
	Print Newspaper	Online Newspaper
Faculty Members	43 (53.25)	37 (46.25)
PhD Students	53 (60.23)	35 (39.77)

(Figures in parentheses denotes percentage)

Table 1 reveals that majority of the faculty member (88.75) and PhD students (95.45) read online newspapers in Aligarh Muslim University (AMU). At the same time when asked about the newspaper preferences, faculty members (53.25) and PhD students (60.23) answered that they prefer print newspaper (Table 2). However 46.25 percent of faculty members and 39.77 percent of students preferred online newspapers. The respondents opined that online newspaper doesn't give the feelings of reading a newspaper.

Table 3: Frequency of online newspaper reading

Frequency of reading	Category of Respondents	
	Faculty Members	PhD Students
Daily	29 (36.26)	32 (36.36)
Weekly	27 (33.75)	17 (19.39)
Occasionally	15 (18.75)	35 (39.77)
On special occasion only	9 (11.25)	4 (4.54)

(Figures in parentheses denotes percentage)

Of the PhD students and faculty members who participated in the survey, majority of the former (39.77) read ONP occasionally. But a 36.26 percent of faculty members and 36.36 percent of PhD students read online daily. A 33.75 percent of faculty members and 19.39 percent PhD are reading weekly and a minority of 11.25 percent faculty members and 4.54 percent students read online on special occasions only. The table depicts that a good number of respondents are reading online newspapers daily.

Table 4: Time spend for reading online newspapers

Time Spend	Category of Respondents	
	Faculty Members	PhD Students
10-30 minutes	39 (48.75)	37 (42.04)
30-45 minutes	34 (42.50)	27 (30.60)
1 hour	7 (8.75)	16 (18.18)
1-2 hour	0 (0.0)	8 (9.09)

(Figures in parentheses denotes percentage)

Of the participant faculty members, 48.75 percent and 42.04 percent of PhD students spend 10-30 minutes to read ONP at a time. A 42.5 percent of faculty members and 30.6 percent of PhD students spend 30-45 minutes to read the same while 18.18 percent students and 8.75 percent faculty members sit one hour for reading ONP. Only a 9.09 percent of students find 1-2 hours to spend on reading online.

Table 5: Preferred language

Language of Newspapers	Category of Respondents	
	Faculty Members	PhD Students
English	52 (65)	52 (59.09)
Hindi	17 (21.25)	20 (22.72)
Urdu	11 (13.75)	10 (11.36)
Other Regional languages	0 (0.0)	6 (6.81)

(Figures in parentheses denotes percentage)

Answering to the question about the preferred language, 65 percent of faculty members and 59.09 percent of PhD students said that English was their preferred language followed by Hindi (21.25 faculty members and 22.72 PhD students) and Urdu language is preferred by 13.75 percent faculty members and 11.36 percent PhD students. Among PhD students, 6.81 percent like other regional languages such as Kashmiri, Malayalam, Tamil, Telugu and Arabic. This group is those from outside Hindi/Urdu belt of India and includes foreign students.

Table 6: Favourite online newspapers

Online newspapers	Category of Respondents	
	Faculty Members	PhD Students
The Hindu	39 (48.75)	52 (59.09)
The Times of India	16 (20)	52 (59.09)
The Hindustan Times	11 (13.75)	12 (13.64)
Indian Express	7 (8.75)	12 (13.64)
Dainik Jagaran	11 (13.75)	22 (25)
Rashtriya Sahara	6 (7.50)	0 (0.0)
Siasath – Urdu	12 (15)	10 (11.36)
Other regional languages	0 (0.0)	20 (22.73)
	Multiple answers were permitted	

(Figures in parentheses denotes percentage)

**Figure 1:** Home page of The Hindu ePaper

Table 6 reveals that 'The Hindu' daily (Figure 1) is the most preferred online newspaper by faculty members (48.75) and PhD students (59.09) followed by 'The Times of India' (20 % Faculty members and 59.09 % PhD students). Other English newspapers like 'The Hindustan Times' (13.75 and 13.64) and 'Indian Express' (8.75 and 13.64) are also popular.

Among the Indian language, Hindi and Urdu papers are selected more. The 'Dainik Jagaran' (Hindi) (Figure 2) is used by 13.75 percent of faculty members and 25 percent of PhD students. But the 'Rashtriya Sahara', a Hindi daily is used

by a 7.5 percent of faculty members only; no students mentioned this newspaper. Among the Urdu newspapers ‘Siasat’ (figure 3) is read by 15 percent of faculty members and 11.36 percent of PhD students.

Among the other languages ‘Greater Kashmir’ (Kashmiri), ‘Madhyamam online’, ‘Manorama Online’ (Malayalam), ‘Dinamani’ (Tamil), and ‘Pratidin’ (Bengali) are mentioned. A 22.73 percent of PhD students are noted as regular readers of these newspapers.



Figure 2: Homepage of ‘Dainik Jagaran’ Hindi Online



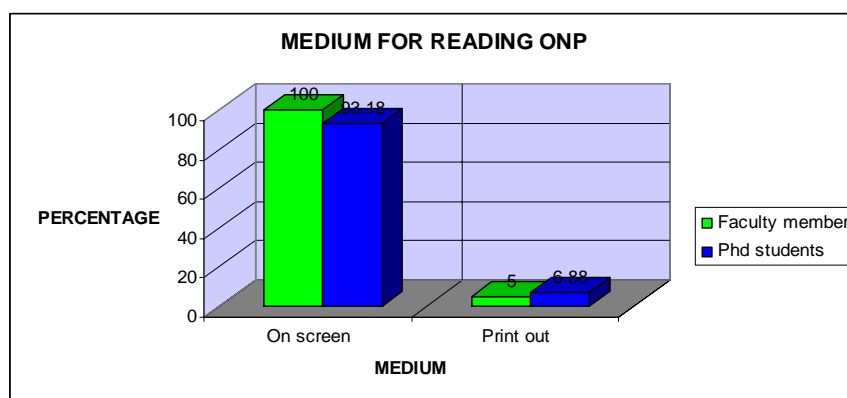
Figure 3: Urdu Online newspaper ‘The Siasat’

Table 7: Reasons for Preference for Online newspapers

Reasons	Category of Respondents	
	Faculty Members	PhD Students
Instant news update	76 (95)	86 (97.72)
Maximum news coverage	50 (62.5)	78 (88.63)
Free of cost	45 (56.25)	72 (81.82)
Searchability	70 (87.5)	63 (71.59)
Easy access to back issues (archives)	62 (77.5)	63 (71.59)
Use of links	30 (37.5)	27 (30.68)
Contact with editors via email	30 (37.5)	27 (30.68)
	Multiple answers were permitted	

(Figures in parentheses denote percentage)

Table 7 reveals the reasons behind the online newspaper preferences among the respondents. Multiple answers were permitted in the questionnaire. It is clear that both faculty members (95) and PhD students (97.72) give first preference to 'instant news update'. A 88.63 percent PhD students said that 'maximum news coverage' is the reason for second preference while faculty members (87.5) feel 'searchability' is the second. In case of faculty members, 'easy access to back issues' (77.5), 'maximum news coverage' (62.5) and 'free of cost' (56.25) are the successive preferences. 'Use of links' (37.5) and 'facility to contact editors via email' (37.5) have equal consideration among them. In case of PhD students, 'free of cost' (81.82) comes in third position; 'searchability' (71.59), 'easy access to back issues' (71.59), 'use of links' (30.68), 'facility to contact editors via e-mail' (30.68) come in successively.

**Figure 4:** Medium for reading online newspapers (ONP)

A question was asked about the medium which is used to read the ONP. 100 percent faculty members and 93.18 percent PhD students read ONP on screen. A minority of faculty members (5%) takes printout occasionally along with online reading and a 6.88 percent PhD students use printout to read ONP (Figure 4).

Table 8: Place of Reading Online newspapers (ONP)

Places where reading ONP	Category of Respondents	
	Faculty Members	PhD Students
Department computer lab	2 (2.5)	33 (37.5)
Central Library (Maulana Azad Library)	6 (7.5)	18 (20.46)
Central Computer lab	5 (6.25)	24 (27.28)
Home / hostel	9 (11.25)	12 (13.64)
Internet café	0 (0.0)	5 (5.68)
Faculty Chamber	58 (72.5)	2 (2.27)
	Multiple answers were permitted	

(Figures in parentheses denotes percentage)

Table 8 shows that majority of faculty members (72.5) read ONP in their own chambers in the departments and a few (7.5) at university central library (Maulana Azad Library) and at their residence (11.25).

PhD students (37.5) rush to the departmental computer lab for reading ONP while at the same time, university central computer lab (27.28), central library (20.46), Hostels (13.64) are also used for this purpose. Interestingly, no faculty members are going to internet café for reading ONP while some students (5.68) read ONP at café also.

The study revealed that the PhD students are not strict about the choice of a particular place to read ONP. An 18.18 percent students use both central library and their respective department lab. Another 12.5 percent students go to department lab, central computer lab and central library for this purpose. But faculty members are not reported to use all these variety of places.

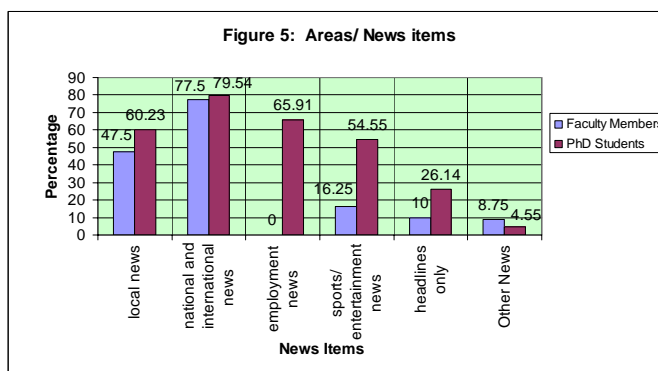


Figure 5: Areas/News items read in online newspapers

A multiple answers question (Figure 5) related to the news items usually read by the respondents reveals that the 'national and international news' attracts 77.5 percent of faculty members and 79.54 percent of PhD students. The figure shows that students go through a variety of news items compared to faculty members. 60.23 percent PhD students and 47.5 percent faculty members read local news. In the case of PhD students 65.91 percent look for employment news but faculty members never go through these columns. Sports/Entertainment news attracts both faculty members (16.25) and PhD students (54.55) and few respondents - faculty members (10 %) and PhD students (26.14) read only headlines. Other news items like political and social commentary are read by a minority of faculty members (8.75) and PhD students (4.55).

Table 9: Use of special feature of online newspapers

Special Features	Category of Respondents							
	Faculty Members				PhD Students			
	Regularly	Frequently	On special occasions	Never	Regularly	frequently	On special occasions	Never
Video Clips	0	9 (11.25)	52 (65)	19 (23.75)	18 (20.45)	33 (37.5)	20 (22.72)	17 (19.32)
Sound Records	0	11 (13.75)	44 (55)	25 (31.25)	4 (4.54)	16 (18.18)	30 (34.09)	38 (43.18)
Online Surveys	0	7 (8.75)	38 (47.5)	35 (43.75)	5 (5.68)	20 (22.72)	28 (31.81)	35 (39.77)
Response through Forums & Chat room	11 (13.75)	13 (16.25)	20 (25)	36 (45)	4 (4.54)	24 (27.27)	20 (22.72)	40 (45.45)
Response to Advertisement	0	0	0	80 (100)	8 (9.09)	20 (22.72)	36 (40.90)	24 (27.27)

(Figures in parentheses denotes percentage)

Table 9 shows the use of special features of online newspapers by the respondents. The comparison between two groups of respondents reveals that PhD students are more interested to use these facilities. Most of the faculty members are using these facilities on special occasions only. Sound records and video clips attached to news items are special attraction of online newspapers; in this case a 65 percent of faculty members watch video clips on special occasions, 11.25 percent of them watch it frequently and a 23.75 percent never. Among the faculty members 13.75 percent frequently listen to sound records attached along with news items and a 55 percent on special occasions only. Online surveys conducted by newspapers on particular events attract faculty members. A 47.5 percent of them participate in online surveys on special occasions while 8.75 percent frequently. Faculty members regularly respond (13.75) to the columns and feature reports through Forums and Chat Rooms when a 16.25 percent respond frequently and a 25 percent on special occasions only. The data reveals that the faculty members (100 %) do not respond to advertisement published in online newspapers.

Among the PhD students 20.45 percent regularly watch video clips and 37.5 percent frequently. 22.72 percent watch on special occasions only while 19.32 percent never watch. In case of sound records attached to news items, 34.09

percent listen on special occasions and 18.18 percent frequently. But a majority (43.18) of them does never listen to sound records. Response to online survey being conducted by online newspapers goes like a 31.81 percent on special occasions and 22.72 frequently participate; 39.77 percent never participate in surveys. When asked about response to columns and news reports through Forums and Chat rooms, 27.27 percent answered that they respond frequently, 22.72 percent on special occasions only and a minority (4.54%) regularly to this item. In case of advertisement published in newspaper, 40.90 percent respond on special occasion, 22.72 percent frequently and only 9.09 percent regularly.

Table 10: Major Difficulties in reading online newspapers

Difficulties	Category of Respondents	
	Faculty Members	PhD Students
Small font size	14 (17.5)	10 (11.36)
Too many links	16 (20)	10 (11.36)
Lack of portability	3 (3.75)	2 (2.27)
Unknown character display for regional language	18 (22.5)	6 (6.82)
Slow downloading	8 (10)	30 (34.09)
Reading on screen is tiring	10 (12.5)	26 (29.55)
Insufficient news report	11 (13.75)	4 (4.54)

(Figures in parentheses denotes percentage)

Table 10 shows the difficulties in reading online newspapers. It is clear that the difficulties faced by the respondents are different and subjective.

The ‘unknown character display for regional languages’ is the main difficulty faced by the faculty members (22.5) when they read online newspapers, ‘Too many links cause confusion’ (20) and it is time wasting. ‘Small font size’ (17.5) which is not able to zoom in/out in case of some news papers make obstacles. ‘Insufficient and incomplete news reports’ (13.75) is also mentioned as a problems. 12.5 percent of faculty members opined that ‘reading on screen is tiring’. ‘Slow down loading rate’ (10) and ‘lack of portability’ (3.75) are the least mentioned difficulty by the faculty members.

Regarding the difficulties of PhD students, ‘Slow down loading speed’ (34.09) is ranked first. ‘Reading on screen is tiring’ (29.55) is coming second. ‘Small font size’ (11.36) and ‘too many links’ (11.36) have been equally ranked. ‘Unknown character display for regional languages’ (6.82), ‘insufficient and broken news reports’ (4.54) and ‘Lack of portability’ (2.27) are also mentioned as difficulties.

Table 11: Start of reading online newspaper

Time	Category of Respondents	
	Faculty Members	PhD Students
Very recently	15 (18.75)	14 (15.91)
Last one year	20 (25)	20 (22.72)
2-4 years	38 (47.5)	50 (56.81)
5-10 years	7 (8.75)	4 (4.54)

(Figures in parentheses denotes percentage)

The study reveals that majority of the respondents, faculty members (47.5) and PhD students (56.81), started reading online newspaper during last 2-4 years. 25 percent of faculty members and 22.72 percent of students started since last one year. A percentage of 18.75 faculty members and 15.91 students started to read very recently and only 8.75 percent faculty members and 4.54 percent students have 5-10 years experience of reading ONP.

Major Findings

Based on the survey and analysis, the investigators reached the following conclusions:

1. Majority of the faculty members (88.75 %) and PhD students (95.45 %) have a habit of reading online newspapers but students are more enthusiastic than faculty members. (Table 1).
2. Even though the study reveals the increasing interest in online newspapers, print newspaper is the preferred medium of reading news among all groups. 53.25% of faculty members and 60.23 % PhD students prefer print newspapers (Table 2).
3. The study reveals that online newspaper is not used as a daily reading material by many respondents. Only 36% of respondents read online newspapers daily. (Table 3).
4. Majority of the respondents are spending 10-30 minutes at a time to read online newspapers, and no faculty member is spending more than one hour for reading online newspapers at a time but a minor percent (9.09) of PhD students do it (Table 4).
5. Even though the AMU is an Urdu speaking campus, it is interesting to know that majority of the respondents, i.e. 65% faculty members and 59.09% of PhD students, prefer English online newspapers followed by Hindi (21.25% faculty members and 22.72% PhD Students) and Urdu (13.75% Faculty members and 11.36% PhD students). A small percent (6.81) of PhD students, those who are from outside Hindi/Urdu belt, prefer their own respective mother tongues like Kashmiri, Telugu, Tamil, Malayalam and Arabic. No faculty member prefers language newspapers other than Hindi or Urdu (Table 5).

6. Among the English newspapers, *The Hindu* is the most preferred online newspaper with a readership of 48.75% faculty members and 59.09% PhD students. Among the Hindi newspapers, *Dainik Jagaran* is the top choice with a readership of 13.75% for faculty members and 25% for PhD students followed by the *Rashtriya Sahara* (7.5% faculty members). The *Siyasath*, an Urdu newspaper is read by 15% faculty members and 11.36% PhD Students (Table 6).
7. Students from distant states who speak languages other than Hindi/Urdu prefer their respective regional languages. About 22.73 percent PhD students read other regional language newspapers (Table 6) like *Madhyamam online*, *Manorama Online* (Malayalam), *Dinamani* (Tamil), and *Pratidin* (Bengali).
8. Both faculty members (95%) and PhD students (97.72%) prefer online newspapers because of the instant news update. A percentage of 88.63 PhD students and 62.5 faculty members believe online newspapers provide 'maximum news coverage'. 'Free of cost' is a prime reason to prefer online newspapers for faculty members (56.25%) and PhD Students (81.89%) (Table 7).
9. Hundred percent faculty members and 93.18 percent PhD Students read online newspapers on screen. A minimum number of faculty members (5%) and PhD Students (6.88) read through printouts (Figure 4).
10. Majority of the faculty members (72.5%) are reading online newspapers in their chambers in respective departments, while some of the PhD students visit department computer labs (37.5%), central library (20.46%) and central computer lab (27.28%) to read online newspapers. No faculty member goes to internet café to read online newspaper, but a 5.68 percent of PhD students said they do so (Table 8).
11. National and international news items are the major attraction of majority of the online newspaper readers (faculty members 77.5% and PhD students 79.54). At the same time a good number of respondents i.e., 47.5% faculty members and 60.23% students prefer local news items also (figure 5).
12. A good percentage of PhD students (65.91) are interested to read employment news but no faculty member looks into this item. Sports and entertainment news attracts 16.25 percent faculty members and 54.55 percent PhD students. Other news stories like those on social and political issues are read by 8.75% faculty members and 4.55% PhD student (Figure 5).
13. Sound records and video clips attached to news items are special enticement of online newspapers.
65 percent of faculty members and 22.72 percent PhD students watch video clips on special occasions (table 9). 55 percent faculty members listen to sound records on special occasions and a 13.75 percent frequently. Among the PhD students 34.09 percent on special occasions and 18.18 percent frequently listen to sound records (table 9).

14. Surveys conducted by online newspapers attract both faculty members and students. Among the respondents, 47.5 percent faculty members and 31.81 percent students participate in surveys on special occasions (table 9).
15. Some faculty members (13.75%) regularly respond to the columns and feature reports through Forums and Chatrooms while 25 percent respond on special occasions only. Among the PhD students, only 4.54 percent respond regularly and 22.72 percent respond on special occasions (table 9).
16. Advertisements, which are highlighted in many online newspapers, on almost every page are the least accepted items. No faculty member responds to advertisements published in online newspapers but 40.90 percent PhD students respond on special occasions (table 9).
17. There is not any common problem reported by respondents while reading online newspapers in AMU. According to faculty members (22.5%) 'unknown character display for regional languages' is the main difficulty in reading online newspapers but in case of PhD students (34.09%) 'slow downloading rate' is the main problem. 'Too many links' cause confusion to 20 percent faculty members and 11.36 percent PhD students. 29.55 percent of faculty members and 12.5 percent of PhD students opined that reading on screen is tiring (table 10).
18. Most of the respondents, i.e. 47.5 percent faculty members and 56.81 percent PhD students started reading online newspaper within last 2-4 years (table 11). A minor group, (faculty members 8.75% and PhD students 4.54%) have 5-10 years online newspaper reading experience (table 11).

SUGGESTIONS AND RECOMMENDATIONS

To publishers:

1. In some newspapers font size is not convenient for easy reading, so there should be facility to increase the font size.
2. Online newspapers have to provide up-to-minute news updates regularly. Special attention may be given to career / employment news.
3. Online newspapers should devote more space to news than to advertisements.
4. Steps may be taken to avoid the character display problems in regional languages. A standard *Intic* script may be followed by regional newspapers.
5. News and reports given in online newspapers should be as sufficient as in their print counterparts.
6. e-Papers may be available at free of cost or at minimum charge to all readers.
7. Sufficient consideration may be given to local news.

To Aligarh Muslim University:

8. Wi-Fi facility should be provided in the entire university campus for easy access to internet.
9. The number of computers with internet connectivity may be increased in the hostels and computer labs.
10. Online newspaper awareness programmes may be conducted at different levels to promote the reading habit.
11. Faster broadband connectivity and better bandwidth should be provided in order to download large amount of data which may include text, graphics, sounds etc.
12. Subscription to e-Paper should be considered to provide full access to the online newspapers throughout the campus.
13. Digital archives of news clippings may be set up in the university library.
14. Facility to take prints of research related news items / articles may be provided in the campus.

CONCLUSION

Online newspapers are among the most visible entities on the Net associated with the longer conceptions of the society and democracy as representatives of the fourth estate but perhaps more importantly as representative of the regional communities (Riley 1998). It becomes an essential part of the daily life of the community. In this study we could only cover the reading habit of faculty members and PhD students of Aligarh Muslim University. Further study among the other sections of the community may be conducted. As online newspaper is easy to use, fast to retrieve, and up-to-minute, it is recommended to accelerate the growth and popularity of this basic information source of the public.

REFERENCES

1. Jeff Sigmund (2009). Newspaper Web Site Audience Increases More Than Ten Percent In First Quarter To 73.3 Million Visitors. <http://www.naa.org/PressCenter/SearchPressReleases/2009/> (Accessed on January 15, 2010)
2. American Journalism Review. [http://ajr.org.](http://ajr.org/) / (Accessed on January 15, 2010)
3. Meyer, E. (1996). All the newspapers that fit. *American Journalism Newslink*. October 29, <http://www.newslink.org/emcol.html>. (Accessed on 10, January, 2010)
4. Online newspaper (n.d). In *Wikipedia*. http://en.wikipedia.org/wiki/Online_newspaper#Online-only_newspapers. (Accessed on 15, January, 2010)
5. Meyer, E. (1998). An unexpectedly wider web for the world's newspapers. *American Journalism Newslink* March 17, <http://www.newslink.org/emcol10.html>. (Accessed on 10, January, 2010)

6. MSNBC, CN Top Global News Sites in March, NY Times Top Paper. http://blog.nielsen.com/nielsenwire/online_mobile/msnbc-and-cnn-top-global-news-sites-in-march/ (Accessed on 8, January, 2010)
7. Arant, David M, Anderson, Quitney Janna. 2001. Newspaper online editors support traditional standards. *Newspaper Research Journal*. Vol. 22. <http://www.questia.com/googleScholar.qst;jsessionid=LhTfLflVQ2X7FrmnyrNprnf11j4GYJHp3pp9vy5TVFZTmW1KJ0Nn!867454298!610724718?docId=5002436011> (Accessed on 16, January, 2010)
8. Riley, Patricia, et.al. (1998). Community or Colony: The case of Online Newspapers and Web. *JCMC* 4(1). <http://jcmc.indiana.edu/vol4/issue1/keough.html> (Accessed on 18, December, 2009)
9. Arlitsch, Kenning, Yapp, L, Edge, Karen. (2003). The Utah Digital Newspapers Project. *D-Lib Magazine*. 9 (3). <http://www.dlib.org/dlib/march03/arlitsch/03arlitsch.html>. (Accessed on 18, December, 2009)
10. Shankar, Udaya (2000). Online newspapers and magazines - aspects & issues. *DRTC Annual Seminar on Electronic Sources of Information. 1-3 March*. (Accessed on 18, December, 2009)
11. Wilson, Brenda. (2007) Factors that Predict Newspaper Reading Habits in College Students" Paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication, The Renaissance, Washington, DC, Aug 08, 2007 http://www.allacademic.com/meta/p202904_index.html (Accessed on 18, December, 2009)
12. Arulraj, R. (2009). News Paper Reading Habit Among the Professional College Staff Members for Various Points of View: a Survey. Filed under *Asian Programming*, November 20. www.assiancooperation.com. (Accessed on 10, January, 2010)
13. Neuberger, Christoph, et.al. (1998). Online--The Future of Newspapers? Germany's Dailies on the World Wide Web. *JCMC* 4 (1) September. <http://jcmc.indiana.edu/vol4/issue1/neuberger.html> (Accessed on 10, January, 2010)

Online newspaper sites

1. The Hindu ePaper (English). http://epaper.thehindu.com/svww_index1.php
2. The Times of India (English). <http://www.timesofindia.com/>
3. The Hindustan Times (English). <http://www.hindustantimes.com/>
4. The New Indian Express (English). <http://www.expressindia.com/>
5. Dainik Jagaran (Hindi). <http://in.jagran.yahoo.com/>
6. Rashtriya Sahara (Hindi). <http://www.rashtriyaSahara.com/>
7. Madhyamam Online (Malayalam). <http://www.madhyamam.com/>
8. 'Malayala Manorama' (Malayalam). <http://www.malayalamanorama.com/>
9. 'Dinamani' (Tamil). <http://www.dinamani.com/>
10. 'Pratidin' (Bengali). <http://www.sangbadpratidin.in/index.php>
11. 'The Siasat Daily' (Urdu). <http://www.siasat.com/>

OPEN ACCESS MODEL FOR LIBRARIES AND NEWSPAPERS: NEW ROLES AND CONVERGENCES

Dr. Ajit Pyati

ABSTRACT

Changing economic patterns and the shift to digital environments have presented a series of challenges both to libraries and the newspaper industry in general. This paper discusses some of the common challenges libraries and newspapers face, and how new synergies between the library and journalism profession might emerge. In particular, libraries have the potential to play new and expanded roles as stewards and access points for digital newspapers. A deeper understanding of the complementary roles libraries and newspapers play as part of the democratic public sphere is needed for new partnerships to take place.

LIBRARIES AND NEWSPAPERS: CHALLENGES OF A CHANGING ENVIRONMENT

The relationship between libraries and newspapers, at first glance, appears to be a simple one. For example, libraries can provide access to newspapers and can also preserve newspapers for archival purposes. This basic relationship based on access and preservation is an important one, but belies some of the other significant relationships between libraries and newspapers. This paper briefly discusses some of the other important connections between libraries and the newspaper industry, with specific reference to changes in economic and technological environments. These changes are affecting both libraries and newspapers in similar ways, and I suggest new collaborative and synergistic opportunities based on the principle of open access and the enhancement of the democratic public sphere. By identifying similar pressures and challenges facing both institutions, new opportunities and partnerships can be realized. The next few paragraphs briefly discuss some of the challenges facing both newspapers and libraries.

The newspaper industry, especially in the Western countries, is facing a number of difficulties. These challenges include declining print circulations, the economic viability of digital newspapers, plummeting advertisement revenues, and major cutbacks of newspaper budgets. For instance, reductions in the number of editorial staff and reporters have hastened the deterioration of several esteemed newspapers such as *The Los Angeles Times* and *The Baltimore Sun*. Some cities in the United States are even at risk of losing their daily newspapers altogether. Along with these challenges is the overarching environment of media consolidation, where a few media companies (e.g., *Time Warner*, *Clear Channel*, *News Corp*, etc.) own a huge proportion of the television networks, radio stations, and newspapers in a number of different media markets. This

oligopoly of media ownership threatens diverse, alternative, grassroots, and oppositional voices.

The declining health of the newspaper industry also is related to the decline of the public sphere of communication in democratic societies. As Jurgen Habermas (1991) has noted, the public sphere is a vital space of dialogue and debate for democratic societies, free from the domination of the state and the market. Institutions such as newspapers and libraries are considered to be prominent parts of the public sphere, where information important to the functioning of vital democracies can be furnished. However, in this age of advanced technological capitalism, the public sphere is shrinking and is subject to the cutthroat logic of profit above all else. In the so-called “information society,” an environment of neo-liberalism and corporate globalization prevails, where global capitalism has greatly extended its reach and is the only game in town. By neo-liberal, what is meant here is increasing privatization and deregulation of public services (Webster, 2004).

Similarly, libraries are also facing a number of challenges in an age of accelerated capitalism and technological change. This environment of neo-liberalism and information commodification manifests itself in issues of copyright and intellectual property for digital resources, declining financial support for public libraries, skyrocketing prices for scholarly journals, and the threat of increasing privatization of library services, to name a few. For example, the extension of copyright and intellectual property laws poses a threat to traditional fair use policies for libraries, and can serve to stifle creativity and increase the concentration of power of ownership in the digital world (Lessig, 1999). In addition, the scholarly publishing “crisis” of rising journal prices has forced many academic libraries to make the difficult decision of cutting subscriptions to many journals (Willinsky, 2006).

Given these various crises affecting both libraries and the newspaper industry, it is fruitful to imagine new types of partnerships between them. Libraries still have an important role to play in providing access to and preserving newspapers; however, we can also find ways of enhancing the roles of libraries and newspapers in the public sphere through partnerships. For instance, we could imagine libraries hosting and providing access to digital community-based newspapers or perhaps providing indexing services and managing access to digital versions of major newspapers. I argue that the concept of open access is important to frame these different types of library-newspaper partnerships. But before I discuss open access, it is useful to understand how libraries are taking on new roles in the publication process.

LIBRARIES, ELECTRONIC PUBLISHING, AND OPEN ACCESS

While libraries in the last several years have transformed themselves into important technology access points, they are also in a position to articulate broader technology goals and development strategies. In response to various economic pressures, for instance, academic libraries are taking on important roles with regard to the scholarly publication process. The development of institutional repositories and exploration of electronic and open access

publishing models have made academic libraries important players in the debate over the future of scholarly publishing (Willinsky, 2006). While academic libraries are not replacing the role of traditional publishers, the role of the academic library in advocating for and supporting technological solutions to support new forms of publishing is significant.

The transformation of the library into an important information technology hub in the age of the Internet is a process that is both old and new. Libraries have been involved in the development of their own technology for many years, with various homegrown systems for library applications developed in the 1970s and 1980s (Morgan, 2002). What makes the new information technology environment in libraries so different, however, is the networked power of the Internet (Morgan, 2002), and the ability to share resources and expertise to better serve the needs of their user communities.

Some of these new roles include an enhanced role for libraries in the publication process, which can encompass tasks as varied as electronic publishing support services, digitization of cultural and academic resource materials, and community archiving, as well as other forms of electronic content management. Libraries have long been in the “content business” — however, rather than merely providing access to content, they are taking more active roles in the development of content itself.

The expanded role of libraries in the publication process did not come by accident — instead the extreme economic pressures of the scholarly publishing environment has been a major impetus. Commercial publishers dominate the scholarly publication market, and have effectively created a bottleneck in the distribution of scholarly information through highly restrictive copyright rules and exorbitant prices for journals. As will be discussed, the enhanced role of academic libraries in the publication process is directly addressing this bottleneck in the distribution of scholarly content.

The severity of this economic situation, the “scholarly publication crisis,” has been one of the defining challenges of the academic library environment over the past decade. The increasing volume and costs of scholarly publications, particularly in science, technology, and medicine, has made it difficult for academic libraries to support the collection needs of their user communities (ARL, 2000). While people outside of the higher education community may not realize the extent of this crisis, it is both shocking and severe, and is a prime example of increasing corporate control over information.

This crisis is based on the successful transformation of knowledge into a capitalized commodity and economic driver (Willinsky, 2006). Major academic journal publishers over the last decade have merged, and the resulting corporate publishing concentration, with its focus on knowledge capitalization and shareholder value, has seen journal prices rise well above inflation rates, and university libraries cannot keep up (Willinsky, 2006). Consequently, many economically less well off libraries and academic institutions have had to massively cut journal subscriptions, with academic libraries in the developing world particularly at a disadvantage.

Given these pressures, the open access movement has taken hold over the last several years. Open access can take many forms, and open access electronic publishing is often cited as an example. Internet technologies have allowed the wide dissemination of scholarly research — allowing libraries, scholars, and publishers alike to re-envision models of scholarly publication. Open access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions, made possible by the Internet and the consent of the author or copyright holder (Suber, 2004). Open access is compatible with peer review and is not free to produce — it is not focused on whether scholarly literature can be made costless, but whether there are better ways to pay the bills than by charging readers and creating access barriers (Suber, 2004). The two most common forms of open access are open access repositories and open access journals. While a topic that has been gaining momentum in recent years, open access represents a growing consciousness around the need to make knowledge and information as widely accessible as possible (Willinsky, 2006). In fact, the very possibility for open access has been greatly enhanced by the presence of digital technologies (Willinsky, 2006).

Libraries have been active on both the open access repository and journal fronts, as academic libraries see the benefit of alternative publication models since they face both a pricing and permissions crisis. The pricing crisis means that libraries must pay increasingly steep prices for journals, while the permissions crisis means that libraries are hamstrung by licensing terms and software locks that prevent them from using electronic journals in the same full and free way as print journals (Suber, 2003). As Peter Suber (2003) argues, librarians can do a lot to alleviate these crises, as they have the best understanding of the problem and can promote open access publishing in their institutions.

Open access has various implications for libraries, and might entail potentially new roles for libraries. For instance, in the mixed open access–traditional publishing environment, entrepreneurial libraries will find new ways to serve their patrons (Schmidt, *et al.*, 2005). In relation to libraries' expanded roles with relation to institutional repositories, another role that libraries may take is to encourage open access publication by subsidizing authors' fees in open access venues (Schmidt, *et al.*, 2005). Funding open access publication provides a new perspective on the library's traditional role as the institutional purchaser of scholarly information (Schmidt, *et al.*, 2005). Thus, the open access movement provides various sets of opportunities and challenges for libraries, including expanded opportunities for libraries to shape the terrain of scholarly publishing.

The involvement of libraries in supporting new forms of publication is a shift in traditional library roles. While libraries are not replacing publishers, supporting repositories and electronic publishing is making libraries important players in the publication process. The case of libraries supporting open access journals is an evolving area. Various commercial electronic publishing software packages exist, as well as some open source software products. A leading open source product is the Open Journal Systems (OJS) journal management and publishing software that was developed by the Public Knowledge Project at the University of British Columbia and is now managed by the Simon Fraser University Library (Willinsky, 2005). An evolving project, OJS has a direct relationship

with an academic library for its technical development, support, and hosting (Willinsky, 2005).

OJS has been a successful open source product, with several hundred journals using this software (Public Knowledge Project, n.d.a). Much of the user base for OJS, in fact, comes from the developing world, with over 200 journals in Africa using the OJS software through the African Journals Online program (Public Knowledge Project, n.d.b). This large amount of uptake in the developing world is not surprising given the economic challenges of accessing commercially controlled scholarly information in that part of the world. The open source nature of the product (free to download) certainly makes it an attractive product for users, as traditional corporate models of scholarly publishing can be bypassed. Defining the library's role in this type of electronic publishing project is a work in progress, but can take various forms.

For example, Willinsky (2006) proposes the idea of a publishing and archiving cooperative among libraries, scholarly societies, and publishing groups. He argues that membership in a publishing and archiving cooperative would enable libraries to participate more directly in journal publishing and archiving to ensure affordable access to research and scholarship. In this cooperative model, libraries would bring their expertise of hosting, indexing, and archiving, while scholarly associations would manage the journal activities — member libraries would pay fees to the cooperative, perhaps based on institution size or on some proportion of the subscription fees they once paid for journals that are now part of the cooperative (Willinsky, 2006). An open source journal publishing software product such as OJS could be utilized in this framework, as well as other electronic publishing software. Since many journals run by scholarly societies are struggling to survive and are moving to commercial publishers for support, perhaps this type of cooperative model might be viable. However, this is just one example of how libraries can take on new roles in transforming scholarly publishing, utilizing the power of Internet technologies to do so.

CONCLUSION – A LIBRARY-NEWSPAPER RELATIONSHIP BASED ON RENEWING THE PUBLIC SPHERE

Now with an understanding of emerging roles for libraries in supporting electronic publishing and the origins of the open access movement, we can see how these ideas might apply to newspapers. For example, we can think of ways in which libraries could host community-based newspapers and magazines. Many of these publications struggle to survive with low circulation numbers, but nonetheless play an important role in providing specific and relevant information for their local communities. Libraries might be able to host an electronic platform for these digital community newspapers, managing user subscriptions, archiving old issues, and providing access. A number of different access models could exist – open access could serve as a guiding principle to make the newspaper's content as widely available as possible, but different payment for access scenarios could be employed. This type of partnership could be a win-win situation for the library and the newspaper, as it highlights the importance of the library to the community and prevents a community newspaper from potentially going under.

In the case of larger newspapers, large and prominent research libraries could enter into partnerships and cooperatives to index and abstract these publications, potentially reducing newspaper operating costs. This could be a collective effort between different large institutions, with some of the subscription payments covering indexing costs. The digital preservation of these newspapers could also take place between these different institutions. Just as *Google* has partnered with certain libraries in book digitization projects, libraries can also partner with newspapers on digitization initiatives. These are just a few speculative possibilities which would need further investigation. Many more fruitful partnerships might potentially exist.

As we have seen from the previous section, a shift in the economic environment of scholarly publishing has forced academic libraries to take an advocacy stance and re-define the nature of scholarly publishing. This expansion of library roles in the publication process is made possible by the transformation of the technological environment and the advancement of the Internet. Libraries might also be in a position to re-define the nature of newspaper publishing. However, a deeper understanding of the complementary roles that libraries and newspapers play in strengthening an embattled public sphere is needed. This understanding can be the first step in developing more fruitful and beneficial collaborations between libraries and newspapers.

REFERENCES

1. Association of Research Libraries (ARL), 2000. "Principles for emerging systems of scholarly publishing," at <http://www.arl.org/resources/pubs/empe/index.shtml>, accessed 24 August 2007.
2. J. Habermas. 1991. *The structural transformation of the public sphere: an inquiry into a category of bourgeois society*. Cambridge, MA: MIT Press.
3. L. Lessig, 1999. *Code and other laws of cyberspace*. New York: Basic Books.
4. E.L. Morgan, 2002. "Possibilities for open source in libraries," *Information Technology and Libraries*, volume 21, number 1, pp. 12–15.
5. Public Knowledge Project, n.d.a. "Current research and development," at <http://pkp.sfu.ca/research>, accessed 30 November 2006.
6. Public Knowledge Project, n.d.b. "Open journal systems — publications," at <http://pkp.sfu.ca/publications>, accessed 30 November 2006.
7. K.D. Schmidt, P. Sennyey, and T.V. Carstens, 2005. "New roles for a changing environment: Implications of open access for libraries," *College & Research Libraries*, volume 66, number 5, pp. 407–416.
8. P. Suber, 2004. "A very brief introduction to open access," at <http://www.earlham.edu/~peters/fos/brief.htm>, accessed 29 September 2006.
9. P. Suber, 2003. "Removing barriers to research: An introduction to open access for librarians," *College & Research Library News*, volume 64,

number 2, at <http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2003/february1/removingbarriers.cfm>, accessed 8 October 2007.

10. F. Webster, 2004. "Introduction," In: F. Webster (editor). *The information society reader*. London: Routledge.
11. J. Willinsky, 2006. *The access principle*. Cambridge, Mass.: MIT Press.
12. J. Willinsky, 2005. "Open journal systems: An example of open source software for journal management and publishing," *Library Hi-Tech*, volume 23, number 4, pp. 504–519.

NEWSPAPER DIGITAL LIBRARIES NEWS-CLIPPING SERVICES AND LONG TERM ARCHIVING USING GREENSTONE

M.G. Sreekumar

R. Biju

T. Sunitha

P. Sreejaya

Joshy Kuriakose

and

K.M. Sudheesh Kumar

ABSTRACT

The importance of newspapers for people of all walks of life, institutions, government and countries in general, need not be overemphasized. It is truly a reflection and narrative of a society, status and its culture. Researchers too turn to newspapers for casual as well as quality research. It is interesting to note the continued relevance of the newspapers in the traditional paper format even when the publishing and transmission of news as well as newspapers across the world have migrated to newer paradigms and platforms such as the Web, Blogs, PDAs, Mobiles, eReaders etc. At the same time, newspaper management in libraries pose plethora of problems to librarians. Requirement of vast space, large size racks, extraordinary weight of bound volumes coupled with the damage that the paper undergoes over time are just a few to mention them. Recent developments in technology, especially the digital technologies, have a lot to offer in efficiently combating the above problems. The current information environment prompts libraries to leverage on the latest digital technologies towards building newspaper digital libraries and in setting up dynamic newspaper access and retrieval systems. These newer breeds of information services offer lots of power and visibility to newspapers, especially the old collections. This paper highlights the power of Greenstone software in developing newspaper digital libraries. It also reports the success stories of two large scale newspaper digital libraries - the Papers Past newspaper digital library, a collection of 19th and 20th century newspapers from the National Library of New Zealand and the Singapore National Library. The experience of the Center for Development of Digital Libraries (CDDL) of IIM Kozhikode in setting up a digital library of newsclippings, the digital equivalent of the traditional press clipping service, using Greenstone software is also shared.

Key Words (newspapers, newspaper digital libraries, newsclipping service, digital library software, Greenstone)

NEWSPAPERS

Newspapers are considered to be the first draft of history, while at the same time, they are part of a country's cultural heritage [Gatos]. Whether it is casual reading, random search or serious research, one of the best places to look for quality information is newspapers. They are rich with local news, long hours of journalism and well researched analysis. For more than 400 years, they have carefully and faithfully chronicled every sphere of life – be it society, politics, education, business or sports. Its value cannot be judged by its enterprise nor its importance derived by its age. Older the newspaper, higher is its research value propositions. Even before computers were born, librarians have developed excellent tools and technologies for storing and retrieving information from the traditional print based collections. The comprehensive indexing and search mechanisms deployed by librarians to retrieve information have been quite accurate and quick. Newspapers are part of the prominent collections of most libraries and the newspaper section is probably the users' first port of land in a library.

Changing Service Models

Even though the manual efforts are commendable, it is simply not enough in an age which is monopolized by search engines and intelligent information systems. Over the years, users also have become very keen in getting instant access to information and they look for micro-second search results. Inefficiencies of manual access, risks of natural hazards and physical storage costs still remain a challenge for the traditional systems to deal with. As technology marches past at tremendous pace, the traditional meaning and definitions of a library's collection range also undergo a great deal of change. In the current information environment libraries need to leverage on the latest digital technologies towards building newspaper digital libraries. By converting newspaper archives to digital collections the dual goal of digital libraries in terms of preventing paper deterioration as well as providing online full-text access of the archives by all interested parties is achieved.

Digital Libraries Carved out of Open Source Software

Digital Libraries (DL) are now emerging as a crucial component of global information infrastructure, adopting the latest information and communication technology. Digital Libraries are networked collections of digital texts, documents, images, sounds, data, software, and many more that are the core of today's Internet and tomorrow's universally accessible digital repositories of all human knowledge. According to the Digital Library Federation (DLF, USA - <http://www.dlf.org>), "Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities".

Digital libraries offer new levels of access to broader audiences of users and new opportunities for library and information science field to advance both

theory and practice. They contain information collections predominantly in digital or electronic form. Electronic publications have some special management requirements as compared to printed document. They include infrastructure, acceptability, access restrictions, readability, standardization, authentication, preservation, copyright, user interface etc.

Digital libraries do enable the seamless integration of the scholarly electronic information, help in creating and maintaining local digital content, and strengthen the mechanisms and capacity of the library's information systems and services. They increase the portability, efficiency of access, flexibility, availability and preservation of digital objects. Once the information is made digital, it could be stored, retrieved, shared, copied and transmitted across distances without having to invest any additional expenditure. Digital Libraries can help move the nation towards realizing the enormously powerful vision of 'anytime, anywhere' access to the best and the latest of human thought and culture, so that no classroom, individual or a society is isolated from knowledge resources. Digital library brings the library to the user, overcoming all geographical barriers.

World over there is increasing appreciation of the Open Access movement and the Open Source Software philosophies and for many a libraries it is a chosen decision, be it technical or financial reasons, not to go for a proprietary digital library software. One needs to evaluate some of the popular Open Source Software for digital libraries, which are in use internationally. 'Dienst', 'Eprints', 'Fedora', 'Greenstone' etc. are among the candidates for the preferred software. Obviously Greenstone outscores the group as a general purpose digital library software from the point of view of a multi-publication type, multi-format, multi-media and a multi-lingual practical digital library [Greenstone]. And once finalized, it could be formally adopted as the software for creating the digital library.

Greenstone Software Features

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. It provides a new way of organizing information and publishing it on the Internet in the form of a fully-searchable, metadata-driven digital library. It has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Its developers received the

2004 IFIP Namur award for "contributions to the awareness of social implications of information technology, and the need for an holistic approach in the use of information technology that takes account of social implications". Again in 2008, Greenstone was bestowed the prestigious Mellon award, which it richly deserved.

Greenstone software along with Java Run Time Environment (JRE), Ghostscript and ImageMagick are deployed for building digital libraries. The software suite is available at the open source directory 'Sourceforge' [Sourceforge].



Figure 1: IIMK Digital Library

The salient features of Greenstone are basically taken from two of the official publications of the software development team appeared in D-Lib Magazine during the year 2001 [Witten, 2001] and 2003 [Witten, 2003]. Greenstone builds collections using almost popular and standard digital formats such as HTML, XML, Word, Post Script, PDF, RTF, JPG, GIF, JPEG, MPEG etc. and many other formats which include audio as well as video. It is provided with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt entirely automatically. The system is extensible: software "plug-ins" accommodate different document and metadata types. Greenstone incorporates an interface that makes it easy for people to create their own library collections. Collections may be built and served locally from the user's own web server, or (given appropriate permissions) remotely on a shared digital library host.

End users can easily build new collections styled after existing ones from material on the Web or from their local files (or both), and collections can be updated and new ones brought on-line at any time. The Greenstone Librarian Interface (GLI) is a Java based GUI interface for easy collection building. Greenstone software runs on a wide variety of platforms such as Windows, Unix / Linux, Apple Mac etc. and provides full-text mirroring, indexing, searching, browsing and metadata extraction. It incorporates an interface that makes it easy for institutions to create their own library collections. Collections could be built and served locally from the user's own web server, or (given appropriate permissions) remotely on a shared digital library host. The other set of features include OAI plug-in (introduced since the 2.40 version) and DCMI compliance, UNICODE based multi-lingual capabilities and a user-friendly multimedia interfacing

[Unicode]. Further more, it is powered by robust indexing systems such as 'Managing Gigabyte' Plus-Plus ('MG' PP) and Lucene. A very interesting feature of Greenstone is its exhaustive set of well documented and articulated manuals (<http://www.greenstone.org/cgi-bin/library?e=p-en-docs-utfZz-8&a=p&p=docs>) such as

'Installer's Guide', 'User's Guide', 'Developer's Guide', and 'From Paper to Collection' a document describing the entire process of creating a digital library collection from paper documents. This includes the scanning and OCR process and the use of the "Organizer". There is one more interesting documentation 'Inside Greenstone Collections' which clarifies most of the trickier parts of using Greenstone, especially dealing with configuration file for the collection in question. The recent introduction of the 'realistic book' feature of Greenstone is very exciting and shall enthrall the users. It provides collection builders the facility to serve the eBooks in the real-life print counterpart format (<http://nzdl.org/Books>). It is truly exciting with its quick, easy-to-use, and responsive page-turning mechanism, and combines the ability to include hyperlinks and animated media.

There are presently two versions of Greenstone going around, Version 2 and 3, and they are generally referred to as Greenstone2 and Greenstone3. The latest in Greenstone2, as on February 2010, is V.2.83 and that of Greenstone3 is V.03. Greenstone2 will be there for some more time, but ultimately Waikato/Greenstone see that Greenstone3 will replace it. IIMK is presently the UNESCO host for the Greenstone Support for South Asia. The web site <http://greenstonesupport.iimk.ac.in> provides a compendium of information on the software. The eList greenstonesupport@iimk.ac.in offers regular online support to around 200 professionals from the South Asia region.

There is a belief that Greenstone software is ideal only for small to medium size digital libraries. Stefan Boddie et. al., colleagues from DL Consulting and the University of Waikato, the founding members of the Greenstone project, have since clarified that Greenstone is now being used to produce large newspaper collections for the National Libraries of New Zealand and Singapore [Stefan].

'Papers Past' Newspaper Digital Library of New Zealand

Papers Past is a Greenstone-based newspaper digital library of the National Library of New Zealand. *Papers Past* contains more than one million pages of digitised New Zealand newspapers and periodicals. The collection covers the years 1839 to 1932 and includes 52 publications from all regions of New Zealand. When fully built, the scanned and OCR'd collection would have around 20 GB of raw text, 2 billion words, with 60 million full-text searchable unique terms. All images will be fully-searchable. The collection would consist of over 6.5 million newspaper articles, each with its own metadata (much of it automatically generated); and the total volume of metadata is 50 GB. Before being built into a digital library collection the metadata is stored in XML format, which occupies around 600 GB, slightly less than 1 MB per newspaper page

[Stefan].

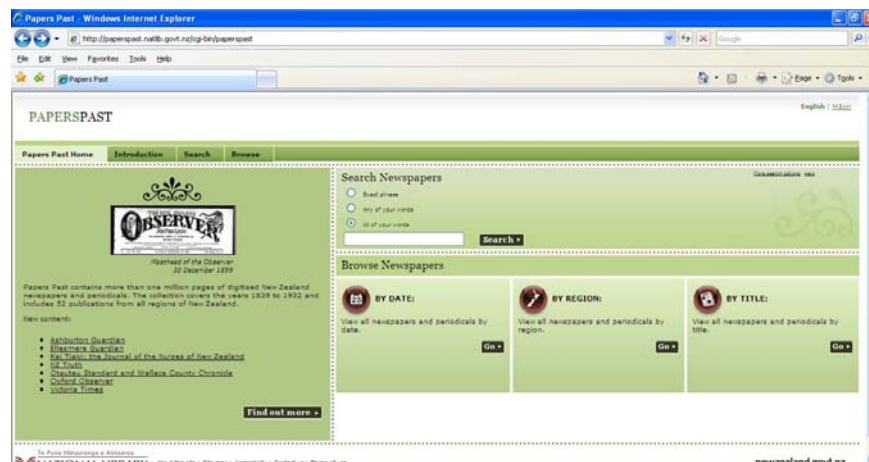


Figure 2: Papers Past Homepage

Papers Past involves a massive amount of metadata. Its specification demands that newspaper articles be viewable individually as well as in their original context on the page. *Papers Past* is presently available at <http://paperspast.natlib.govt.nz>. There are two main ways to find information in *Papers Past*. The searching mode let users enter a query term and retrieves articles that contain that term. The browsing feature let users look at all the newspapers, starting with a year, a region, or a newspaper title. All the newspaper titles on the site can be searched and browsed.

Newspaper Digital Libraries of Singapore National Library

The newspapers digital library at the National Library Singapore is again powered by a Greenstone-based commercial product named Veridian. The front-end user interface as well as the retrieval engine are developed by the National Library Board. The collection is a vast one, comprising around 200 Singapore and Malaya titles published since 1806. It is rich in content and is quite important one, as it reflects the political and social life of Singapore since the country's founding in 1819. English language titles, representing the official language of Singapore, as well as other prominent languages of the ethnic groups in the country such as Malay, Mandarin and Tamil are also there in the collection. The collection is hosted at the Lee Kong Chian Reference Library in the National Library Building. The project is a joint initiative of the National Library Board (NLB)'s National Library and the island-wide network of Public Libraries.

The service, named NewspaperSG, was released on the Web in March 2009. Both onsite and offsite users can search over 550,000 pages of the digitised Straits Times (1845 –

1989) and the microfilm holdings of the newspapers from a single website at <http://newspapers.nl.sg>. Users will also be able to obtain some brief information such as the article title, date of publication, and a 50-words extract from the search results. Full article views are, however, only available through the library's multimedia stations (PC stations). To aid the discovery

of archival news content from the Web, a "Table of Contents" is also generated for each newspaper issue and then submitted to major search engines (Google, Yahoo, MSN) for indexing.

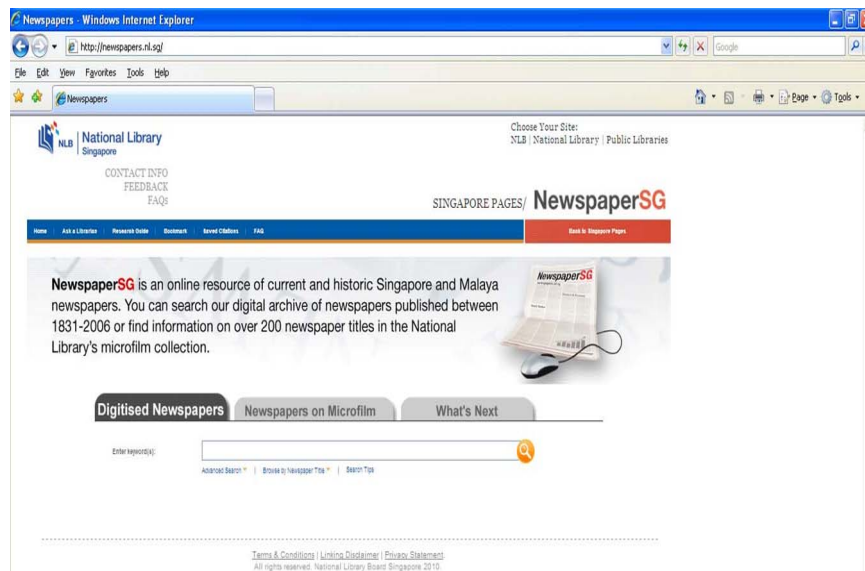


Figure 3: NewspaperSG Homepage

Technical Considerations

Newspapers are pretty complex objects to deal with and they indeed pose a host of challenges to the digital library management and the collection development team. Firstly, in most cases, they manifest in A2 size and that itself is a point of debate as to what should be the mode of presentation in terms of size, in the digital interface. There are multiple service models found in usage by different online newspaper libraries such as ePaper format bundled with proprietary eReaders, PDF/JPEG/TIFF files of different sizes and resolutions or in any other suitable image format. While dealing with quite old and backdated newspapers, the low quality of originals and the old font styles could give substantial labour. OCRing these documents could be even more a mammoth task.

Search tools should be carefully designed, keeping in mind the different category of users. Provision should be there for advanced searching using intelligent algorithms, exact word searching and Boolean queries, while the basic free text searches on simple as well as compound terms are well addressed. Vector space approach, extended Boolean models and weighting schemes are other possibilities for increased search and retrieval efficiency. For non-English and vernacular collections, NLP and its advanced models could be explored and implemented.

Metadata which are searchable by the available indexing systems for each and every article, forming part of the newspaper items in the collection, is the next important challenge. Systems like Greenstone has mechanisms to

generate automatically extracted metadata from most of the text oriented file formats. Bi-tonal digital images in TIFF format are the one recommended for scanned versions and then they are later OCRed for indexing purpose. While Dublin Core, ONIX, METS and ALTO (XML Schema for technical metadata used with OCR scanning output) are the recommended metadata standards, XML is the recommended encoding or representation markup format.

Newsclipping Service of IIM Kozhikode

Inspired by the seamless features of Greenstone software in terms of interface presentation, search and retrieval, customization and configurations, we embarked on creating a digital library of newsclippings. Traditionally, pressclipping services have long been well known artifacts offered by the library fraternity. Named as 'IIMK in Media', the collection aims at containing all news items on IIMK that have appeared in newspapers across the world. A reverse chronology collection development approach has been adopted presently, while parallel efforts will be tried for capturing the news archives available in print format.

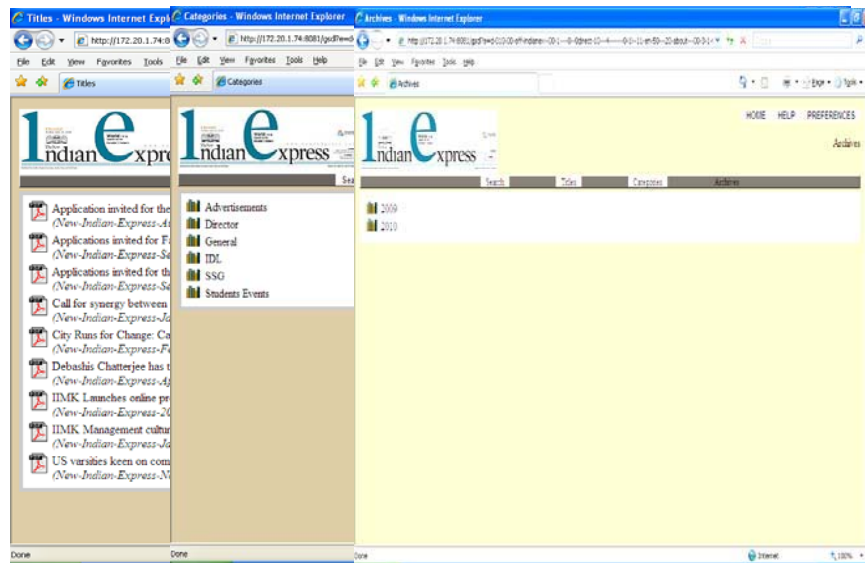


Figure 4: News-clipping Service Homepage

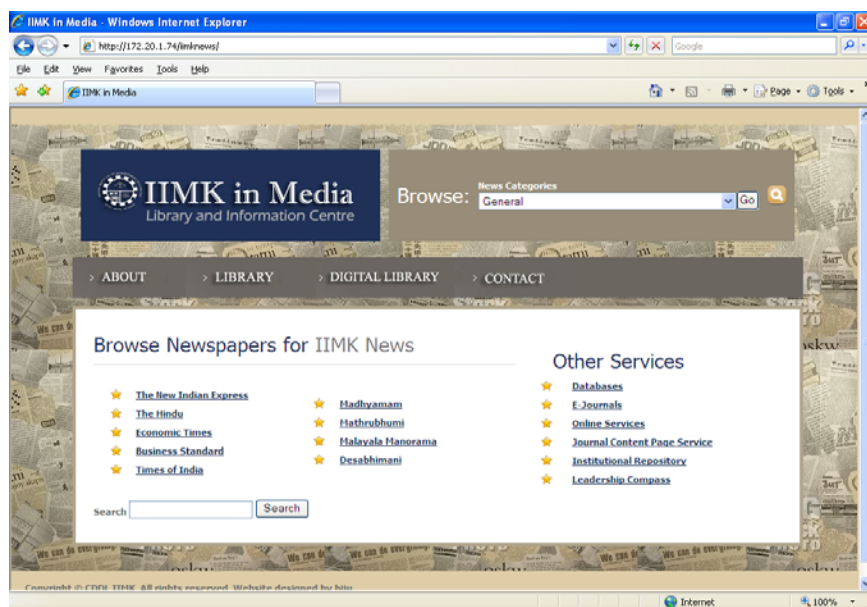


Figure 5: Browsing Classifier Screens : Title, Categories and Archives

The information model chosen for this service at the front-end interface include, in addition to the generic Browse and Search, a pull-down menu of News Categories, and the Newspaper Title-based collections. The collection level (within each newspaper title) browsing classifiers include Titles (the news headings), Categories (Advertisements, General, Programmes, Students, Faculty etc.) and Archives (year-wise). Hierarchy metadata descriptions, using Dublin Core, facilitate browsing on multi-layer subjects, categories as well as years and dates. The search interface could be configured for single collection, multiple collections or searches at the global level.

CONCLUSION

Newspapers play commendable roles in the socio-cultural-political fabric of a society, and preserving them as part of our heritage is of paramount importance. Maintaining them in the traditional format and methods are not viable any more. Digital libraries offer enormous opportunities, options, convenience, flexibility and improved efficiency to the users. Availability of robust software in the open source domain provides an unprecedented opportunity to libraries across the world, especially in catching up with the latest trends and technology. Greenstone provides a vast amount of freedom to the users to leverage on, and being a componentization model, it has the advantage of experimenting with the emerging and cutting edge applications and getting them incorporated into it. The model newspaper digital libraries showcased in this paper are just two examples and one can see more such examples in the Greenstone example collections as well as elsewhere. Also, the newsclipping service mentioned is simply one among such many applications one could think of configuring using Greenstone. The message

that we wish to underscore in this paper is that there are a whole lot of powerful open source software solutions that are built on latest software architectural standards and technologies, and the information science fraternity may find, evaluate, choose and use them as appropriate.

ACKNOWLEDGEMENT

The authors would like to acknowledge the unstinted support extended by Stefan Boddie of DL Consulting, New Zealand, and the development team of the Papers Past the NLB collections.

REFERENCES

1. Gatos, B. et. al.
2. An integrated system for creating a Digital Library from Newspaper Archives
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.282&rep=rep1&type=pdf>
3. 2. <http://www.dlf.org>
4. IIMK Digital Library < <http://iimk.ac.in/gsdll/cgi-bin/library> >
5. 4. <http://www.greenstone.org>
6. SourceForge.net (world's largest Open Source software development website)
7. <<http://www.sourceforge.net/>>
8. Witten, Ian H. et al. 2001
9. Greenstone : Open-Source Digital Library Software
10. *D-Lib Magazine*, 7 (10): 1-16.
11. 7. Witten, Ian H. 2003
12. Examples of Practical Digital Libraries : Collections Built Internationally Using
13. Greenstone
14. *D-Lib Magazine* 9 (3): 1-15.
15. 8. Stefan, B et. al.
16. Coping with very large digital collections using Greenstone
http://www.google.co.in/url?q=http://www.dlconsulting.com/marketing/vldl_boddie_et_al_final.pdf
17. 9. <http://paperspast.natlib.govt.nz>
18. 10. <http://newspapers.nl.sg>

MEDIA ARCHIVES-CUM-NATIONAL REFERENCE LIBRARY ON THE NORTH-EAST (INDIA)

Prof Alaka Buragohain

Keywords: Heritage; Conservation; Digitization; Historical information; Global accessibility; Media Archives; Media Trust; National Reference Library; Newspapers; North-East; Search ability

INTRODUCTION

Newspapers and popular magazines are the mass media of communication. Newspapers are indispensable companion, part and parcel of one's daily life. They are communicating day-to-day events to the public. As such, they are the invaluable testimony to documented stratified knowledge of social change and social history, cultural heritages of the respective people and the places. Newspapers are the first draft of history, chronicles of social, cultural and political past; serving the public as watchdog, investigating the public affairs. Thus, newspaper is the Fourth Estate of a democratic nation and state.

There is no substitute for newspapers primarily for the local media materials which deserve to be preserved and conserved permanently and so to be archived. Contents of the newspapers are the raw materials, may serve as primary sources of historical information provided that the sources of reports are authentic, impartial and reliable.

Popular magazines carry forward news and views of the public, both as primary and secondary information.

MEDIA ARCHIVES-CUM-NATIONAL REFERENCE LIBRARY ON THE NORTH-EAST

INTRODUCTION

A national seminar was organized on 'Social role of media and Building a National Information Resource Centre-cum-Archives for the North-East' during 21-22 September 2004 at Guwahati, sponsored by RRRLF (Raja Ram Mohun Roy Library Foundation), Kolkata, Dept of Culture, Govt of India, organized by Media Trust, Assam. The national seminar inaugurated by Chief Minister of Assam Sri Tarun Gogoi was attended by distinguished journalists and professionals. One of the resolutions adopted in the seminar strongly recommended the establishment of a 'Media Archives-cum-National Reference Library on the North-East' as an autonomous public institution for scientific preservation and servicing of the sources. And Media Trust, Assam that has already made proposals along lines was entrusted to take appropriate steps to implement the recommendations and take up with the concerned authorities for formation of an expert committee to prepare a comprehensive proposal taking

note of all the aspects of objectives, collection, preservation, operations, services and building plan' (Buragohain 2005: 27).

Media Trust, Assam set up in 2002 as the successor of the '150 Years of Newspapers in Assam, Celebration Committee' formed in 1992 to commemorate 150 years of newspapers in Assam since publication of the first Assamese newsmagazine *Orunodoi* in 1848 from Sibsagar. The Trust drafted a project proposal on 'Media Archives-cum-National Reference Library on the North-East' with an estimated cost of about two hundred million rupees for creation of the flexible composite building and the infrastructure.

OBJECTIVE

Under 6th Clause of Assam Accord: The Trust submitted in 2006 the project proposal for institution of the 'Media Archives-cum-National Reference Library on the North-East' with all other paraphernalia to the Govt of India through AASU (All Assam Students' Union) under the 6th clause of the Assam Accord signed in 1985 between the Govt of India and the AASU (Buragohain 2006: 4).

The project has been planned under the provision of Clause No. 6 of the Accord, i.e. 'Constitutional, Legislative and Administrative safeguard as may be appropriate, shall be provided to protect, preserve and promote the cultural, social, linguistic identity and heritage of the Assamese people' in the form of local newspapers, selected local magazines that carry the social and cultural past and also the documents related to different social, cultural, political aspects of the North-East. The Trust considers that the cultural, social, linguistic identity and heritage of the Assamese people imply the culture, the societies, languages and heritage of the indigenous local people of Assam including all the indigenous ethnic groups. The domicile groups also are included in the concept of Assamese people who have accepted and developed the culture and language of the indigenous people in lieu of their own, may be in a modified way.

The Archives will try to restore the past, depict the present and preserve and conserve the national heritage in the form of media materials in best available form for the future generation. Printed materials, at least some specimen copies of the printed media materials will be kept conserved physically as far as possible, the rest will be converted into digitized formats, i.e digitized images but with no audio, animation of multi-media elements.

Official formalities have been completed but still final approval from the Govt of India is awaited. However, collection, processing and partial construction works have been started under different sponsorship.

Healthy Growth of Mass Communication and Media: The other objective is to support a complete and healthy growth of mass communication and media in Assam and facilitate study and research on different aspects of mass communication and media in the region.

Database on the North-East: Further, to create a database for the North-Eastern Region in order to facilitate all decision-making processes with

accurate, relevant and exhaustive information and to support effective journalism with proper information.

Mass Media Vs. Social Integration: Mass media has a very unique role in moulding the society, polity and the government in the right desired direction. In spite of unique diversities among different indigenous people of this region, there is a strong bond of inherent cohesion with common basic ingredients like those in food, in dress, in folk tales, in values, etc which can be explored for moulding the social and political thought for cultural assimilation and social integration in a natural process for the greater national interest without sacrificing the individual identities.

Global accessibility and searchability: The ultimate objective is to make the digital section globally accessible and searchable specifically, exhaustively and instantly and also to obtain the desired materials either by downloading from the Internet or by e-mail from any part of the world.

Information Communication Technology: Application of ICT along with its latest developments is essential in order to utilize its services and facilities and make the Archives a modern one capable of satisfying the present needs and demands.

COVERAGE

Media Archives: The project has been planned to restore, preserve and conserve all the local newspapers and selected magazines in all languages of the state and the rare and important documents and manuscripts related to journalism, media persons of Assam in particular. Retrospective issues will be in original physical documentary form as far as possible and also in converted micro-form, i.e. in CDs, DVDs both offline and on-line.

Moreover, motion pictures, TV and Radio serials also are planned to be preserved and conserved under an air-conditioned environment.

Further, eminent Assamese writers' (both media people and other writers) original manuscripts and handwritings deserve to be preserved. Reportedly it is learnt that one of the original manuscripts of Srimanta Sankardeva (1449-1568) in his own handwriting is in the way of destruction in Madhupur Satra in Koch Behar (now in West Bengal, India).

It is regretted that lots of invaluable materials of our glorious past including different issues of newspapers and periodical literature and other linguistic, literary, cultural and academic heritage have already been lost due to non-availability of scientific facilities for preservation and conservation of them in any national centre.

Current Newspapers and Selected Journals: Current issues of newspapers, news magazines and selected magazines are kept displayed for regular use in a separate section.

Oral Collection and Record: Audio-visual and multi-media cassettes and CDs recording traditional knowledge of illiterate and semi-literate old persons along with their transcripts may be maintained and preserved as an oral collection on the line of British Library in U K.

National Reference Library on the North-East: The Press along with other Mass media as the Fourth Estate of democracy is to inspect, influence, shape and indirectly govern public affairs in a democratic government set-up. It is the most powerful instrument and so ought to be most responsible for shaping public opinion on any issue of importance, a great mediator between public and the government and defender of liberty and rights of the people, addressing always to large audience, to every conceivable group in the society.

The mass media is to entertain, inform, educate and influence by selling all types of information including everything that is of interest to the groups of different categories of the society. As such, whatever is communicated through the mass media requires to be properly informative, accurate, authentic, reliable and in best presentation in order to ensure the effectiveness of the mass communicated messages to the people.

In view of the above fact that the scope of the Press and the Mass media is as vast as the entire universe of knowledge, the project would require also to build up a collection dealing with the entire universe of knowledge in relation to the North-Eastern India in particular. Therefore, a special section as a 'National Reference Library on the North-East' is being created which besides supporting the journalists in different ways would become of special interest to the tourists, area-specialists, anthropologists, ethnologists, historians, planners, decision-makers, diplomats, industrialists, etc.

This project will provide for a national cultural and academic hub presenting its people's variety of ethnic identity with the respective distinct socio-cultural identities, traditions and education resulting in great socio-cultural and socio-economic variations with a large number of dialects and languages.

Besides, physically the North-East is identified as the hotspots of bio-diversity and is virtually a virgin territory in terms of study and research upon enormous natural, forest and water resources.

INFRASTRUCTURAL PROMOTION

Administrative Block: It has been constructed with MPLAD fund of Hon'ble Prime Minister Dr Man Mohan Singh and inaugurated on 15 June 2008 by Shri Tulsi Gobida Baruah, Chairman, Assam Tribune Groups of Newspapers (Buragohain 2008: 14). Part of the Archives has been accommodated there temporarily. The rest has been accommodated in the Journalists' Guest House.

Exhibition Room, Reading Room, Stack, etc: Rare materials of special interest, specimen copies of newspapers and magazines would be displayed for reference under glass/celluloid boxes and also other general exhibits for respective special purposes in separate exhibition room. Reading rooms along with different sections, various stack rooms are essential.

Conservation Laboratory: Newspapers are companions of people in the morning, but in the evening these are discarded as refuses. Newsprints are very fragile, and destroyed very soon. On top of it Assam's climate is humid which is more damaging to newsprints. As such, newsprints require both chemical and physical curative conservation like deacidification, fumigation and are restored by lamination, binding, etc

Digital and Microfilming Unit: Reproductive conservation is a must as newspapers become very bulky and unwieldy with daily additions and so unmanageable. This bulk must be converted into micro form. Besides, media materials must be made globally accessible and at the same time make searchable inside and outside. On-line net-working under a consortium for resource sharing, accessibility of local resources to outside users through Internet as well as conservation of cultural and historical documentary heritage for the posterity besides solving space problem are desirable. Selected ones may be micro-filmed for permanent preservation. For all these an infrastructure with scanner, compatible RDBMS software for specific retrieval of content, microfilming infrastructure, computer, server, etc are essential.

Digital Section created with partial infrastructure was inaugurated on 23 June 2006 by then Chairman and Managing Director of Oil India Shri Mulkh Raj Pasrija (Buragohain 2007: 18)

Documentation-cum-Computer Section: Most of the processing works including documentation works, creation of databases, publication of News digest, Index, etc should be done.

Auditorium and Seminar Halls: Auditorium and seminar halls are essential for holding different seminars and conferences even for outside parties on charges as revenue earning sources.

Research cubicles: Research cubicles are necessary for scholars doing any study and research on any topic using the collection inside and outside of the Media Archives.

Institute of Mass Communication and Media: An Institute of Mass Communication and Media as an essential component of the project has already been instituted since it was inaugurated in 1988 by then Director of the Press Trust of India Sri Ajit Bhattacharya. Study and research on mass communication and media are desirable objectives.

MARKETING MEANS

Collection: Several individuals and families have donated their Preserved retrospective volumes of several local newspapers and magazines like the only available copy of *The Times of Assam*, dated June 15, 1929 (the first English news magazine published weekly since 1888 but no other copy is found available anywhere); *Ajir Asam* (Assamese) starting in 1987 but available from 1993 with number of missing volumes; *Dainik Janambhumi* (Assamese) available since its beginning in 1972 but missing the volume of 1977 and

missing some copies in different years; *Dainik Asam* (Assamese) started in 1965 but available since 1980 but missing some volumes and copies in different years; *Deka Asam* (Assamese) available since its starting in 1935 but missing some copies in different years, etc.

Current issues of most of the local dailies and weeklies and also selected local magazines in local languages amounting to more than thirty since 2003 are collected on regular basis mostly freely and a few on subscription are preserved on regular basis as bound volumes.

Books and documents related to different aspects of the North-East also are donated by a few individuals; other few have donated fund for acquiring books on the North-East.

Processing: Two bibliographical databases have been initiated, one for periodicals, the other for books with Unesco's UNISIS software that is very powerful for information retrieval. Periodicals Database has indicated the available volumes and also information about missing volumes and dates and damaged ones. Books database is indicating the concepts/ keywords occurring in the respective books.

North-East Database, 1996, sponsored earlier by North-Eastern Council has been installed and planned to be up-date, volume by volume. It consists of three volumes, one for titles of periodicals, basic information sources, rare books and manuscripts from the North-East, the second one is for books and the third one is on periodical literature, all related to different aspects of the North-East,

Preservation: Rare and specimen printed copies are to be preserved and conserved as long as possible in its original physical format and the media materials are to be converted into micro form for permanent preservation for the posterity. As such, a National workshop-cum-Training was held during 21-24 November 2007 on 'Preservation and digital Conservation Methods of Newspapers, Magazines and Archival Materials with reference to the North-East', sponsored by National Library and RRRLF, both in Kolkata under the Dept of Culture, Govt of India. Two of its recommendations state 'necessary arrangements be made to preserve and conserve the specimen copies of local newspapers and selected magazines as long as possible' and 'recognizing the importance of microforms of newspapers for long term preservation and the need of digitization to support effective search, global accessibility and retrieval, the workshops recommended that efforts be initiated by Media Trust, Assam to employ appropriate techniques for this purpose' (Buragohain 2008: 18).

An Intensive Training Course for Physical and Chemical Treatment of Printed Rare and Old Documents in the Humid Climate of Assam' would be held shortly sponsored by National Library, Kolkata.

It is decided that digitization of media materials would be done first directly from the printed pages for better image-creation for the purpose of search and retrieval. Only selected materials would be microfilmed later for permanent reservation.

Digitization : Digitization is presently considered a desirable technique in order to achieve our objectives of promotion and marketing of media materials. Selection of a suitable scanner for digitization of the unwieldy newspapers is one of the most crucial decision. However, Bookeye 3 A2 Colour and portable one (German), overhead scanner seems to be the one which may be purchased from a dealer based in New Delhi. Some of the materials will be in on-line, the rest will be in offline in the form of CDs and DVDs which may be sold out for revenue generation.

There will not be much problem for the English prints as OCR (Optical Character Recognition Software) is applicable to the English word images. However, OCR is not applicable to the local languages till now, although some researches are carried on in a local university in this direction. Otherwise creation of Metadata manually with record for every news-item in respective pages of respective newspapers would be much time-consuming.

A project has been planned presently to digitize a local weekly titled *Dhansiri* since 1980 on sponsorship of its management and to create some CDs as per expected demands to be sold out also. WNISIS software has been used to make first the database creating records for each of the items in the respective pages of the weekly under page level and news item level metadata. News-item level will be divided under categorical divisions then under subject wise. Retrieval will be done accordingly and CDs will be made.

CONCLUSION

This project is a little different from simply newspapers libraries. It is still on the footing only. The infrastructure for on-line networking is yet to be completed and to create digital surrogates for the printed materials. Two professionals were deputed for two-months technical internship in Centre for Studies in Social Sciences, Kolkata during July-August 2009 as a first step for joining a Consortium in Social Science and Humanities in India. Experiences obtained in various fields would help in developing this new venture.

REFERENCES

1. Buragohain, Alaka. 2005. A Report on the Media Archives-cum- National Reference Library on the North-East. *The New Spectrum*. V 1 (!): 21-29
2. Buragohain, Alaka. 2006. Digital Section of the Media Archives-cum- National Reference Library on the North-East. *The New Spectrum*. V 2 (!): 3-4
3. Buragohain, Alaka. 2007. Media Archives-cum- National Reference Library on the North-East on the footing. *The New Spectrum*. V 3 (!): 17-18
4. Buragohain, Alaka. 2008. A Report on the Media Archives-cum- National Reference Library on the North-East. *The New Spectrum*. V 4 (!): 14
5. Buragohain, Alaka. 2008. Report on the National Workshop-cum-Training on Preservation and Digital Conservation Methods of Newspapers, Magazines and Archival Materials with reference to the North-East. *The New Spectrum*. V 4 (1): 17-18

ONLINE BENGALI NEWSPAPERS: A COMPARATIVE STUDY

Ashis Biswas

and

Mayuri Das Biswas

ABSTRACT

The number of online newspapers has increased tremendously in the first decade of 21st century and so has their usage. Scholarly and practical thinking about the significance of this new phenomenon has gone through the inevitable stages of euphoria and doom. The Bengali newspapers in India are also not untouched by this latest phenomenon. It is only in the last few years that Bengali newspapers have entered a more temperate period, when publishers are gradually working towards finding the appropriate place for online version of newspapers in news market. The time has come for comprehensive assessment of how online Bengali newspapers are changing newspaper markets. This study discusses the impact of online Bengali newspapers on news and newspaper markets in the Bengali speaking region of India and on the Bengalis living abroad. It also discusses the meaning, history and characteristics of online newspapers. It compares the different online Bengali newspapers in India on different aspects along with advantages and limitations.

Keywords: • Bengali newspaper • Bengali newspaper market • Bengali website • Online Bengali newspaper • Online version • Online edition • Web newspaper

INTRODUCTION

Information is the sine qua for all activities in today's information society. Current information is most demanded and needed by various research groups and individuals. Newspapers are the most prominent sources of current information. They bring to the readers the global information by overcoming all geographical limits. In this regard the online version of newspaper is more helpful. Online newspapers contribute a lot to the 'Information Democracy' and 'Right to Information', since users can access them irrespective of their location. These online editions fulfill the law, 'Information for all' by their universal availability and access.

An online newspaper, also known as a web newspaper, is a newspaper that exists on the World Wide Web or Internet, either separately or as an online version of a printed periodical. Online newspapers are much like hard-copy newspapers and have the same legal boundaries, such as laws regarding libel, privacy and copyright, which also apply to online publications in most countries,

like in the UK. In India the online newspapers are also published following all the legal issues. (http://en.wikipedia.org/wiki/Online_newspaper)

In 1995 many of the newspapers in Europe and US began to publish their online editions. Editors of online medium considered that many political, social and sport events could be published online quickly, competing with TV and radio. Regarding the online Bengali newspapers in India, *Sangbad Pratidin* has launched the first Bengali website in Bengali. www.sangbadpratidin.com.

OBJECTIVES

- To locate the readers of Indian online Bengali newspapers,
- To compare the online Bengali newspapers from different perspectives,
- To identify the impact of online Bengali newspapers on Bengali newspaper market,
- To identify the advantages and limitations of online Bengali newspapers.

READERS OF ONLINE BENGALI NEWSPAPERS

Bengali is the national and official language of Bangladesh and one of the 23 national languages recognized by the Union of India. It is the official language of the state of West Bengal and the co-official language of Tripura, Cachar District of southern Assam and the Union Territory of Andaman and Nicobar Islands. Bengali speakers make the majority in Neil Island and Havelock Islands. It was made an official language of Sierra Leone in order to honour the Bangladeshi peacekeeping force from the United Nations stationed there. It is also the co-official language of Assam, which has three predominantly Sylheti speaking districts of southern Assam – Silchar, Karimganj and Hailakandi. At the international level, Bangla diaspora is found in the countries like Bangladesh (10, 00, 00, 000 -1994), Nepal (27, 712), Singapore (600), United Arab Emirates (70,000), United States of America, United Kingdom, Saudi Arabia (15, 600), Malawi, Canada, Australia, France, etc. (www.lisindia.net/Bengali?beng_Speech.html)

About 67% of Bengalis who use Internet read online Bengali newspapers.

The market segment:

The majority of readership is in the age group of 21-40 years. Professionals in the IT, education and media fields constitute almost half (48%) the readership.

A large number of nonresident Indian (NRI) Bengalis living abroad read online Bengali newspapers.

Proportion of female readers is considerably smaller than male readers – only 13%.

(www.pluggd.in/india-digital-media/online-digital-newspaper-readership-in-india-14-use-it-as-a-tool-to-overcome-homesickness-574)

SCOPE AND COVERAGE

To conduct this study of online Bengali newspapers 11 online Bengali newspapers have been taken into consideration. They are *Aajkaal* (www.aajkaal.net), *Abasar* (www.abasar.net), *Anandabazar Patrika* (www.anandabazar.com), *Bartaman* (www.bartamanpatrika.com), *Daily Desher Katha* (www.dailydesherkatha.com), *Ganashakti* (www.ganashakti.com), *Goldnews* (www.goldnewsindia.com), *Samayik Prasanga* (http://samayikprasanga.karimganj.com), *Sangbad Pratidin* (www.sangbadpratidin.net), *Century Sangabad* (www.suprovat.com) and *Uttaranga Sambad* (www.uttarbangasambad).

COMPARATIVE ANALYSIS

Publication and Distribution

Aajkaal is one of the leading Bengali newspapers in Kolkata. The newspaper was started in 1981 by Abhik Kumar Ghosh, and was part of the transformation of the Indian newspaper industry in the 1980s. The newspaper is known for its left-leaning coverage and its comprehensive sports news. Its readership was measured as 3.3 million by Round 1 of the Indian Readership Survey (IRS) 2008. In the beginning of 21st century it has also started its online version to make it available globally free of cost.

Abasar was launched in 2002 and was created for Bengalis, who are not conversant in English, to have access to information that affects their lives. The name abasar.net comes from the Bengali *abasar*, which means time off. The development of the website was started with the idea that volunteers who had taken time off work could help build this database that would eventually help those in villages of West Bengal and Bangladesh who went online. The website has a total of 20, 000 + web pages. It has a strong focus on women. Among other things it includes information on human trafficking, laws protecting women's rights in India and Bangladesh and tips for law enforcement officers to understand trafficking patterns. The present editor of *Abasar* is Sujan Dashgupta. A committee of advisors comprised of Shankar Sen (President), S. K. Roy, Jayanta Kumar Mitra, Mahashweta Devi, Shamita Dashgupta, Saroj Gupta, Saroj Ghosh, Sunil Gangyopadhyay, Sumit Roy and Tarek Ali look after the various news sections of *Abasar*.

Anandabazar Patrika is a Bengali language newspaper published from Kolkata, New Delhi and Mumbai by ABP Pvt. Ltd. (an enterprise of Ananda Publishers). With 1,277,000 daily copies, it has the largest circulation for a single-edition, regional language newspaper in India. It is the most circulated Bengali newspaper in India. Other than Kolkata, it is also printed from various other towns in West Bengal. The paper was founded in 1922 by its inaugural editor Prafulla Chandra Sarkar. Presently, the newspaper is edited by Aweek Sarkar. The online version of *Anadabazar Patrika* was started from 1 April, 2001.

Bartaman is a Bengali language newspaper published from Kolkata by Bartaman Pvt. Ltd. Apart from the Kolkata edition, the newspaper has four other simultaneous editions, published daily from four major towns of West Bengal -

Siliguri, Burdwan, Malda and Midnapore. The newspaper was founded by Barun Sengupta (1934-2008), a journalist of Calcutta in December 1984. Since April 2002, Bartaman Patrika has published an online version.

Daily Desher Katha which started its publication from 15 August, 1979 from Agartala, Tripura has come a long way to reach the present status. It is the official mouthpiece of the Communist party of India (Marxist) Tripura State Committee. It began its online version at the middle of this decade.

Ganashakti is the official mouthpiece of the Communist Party of India (Marxist) West Bengal State Committee, which first appeared as a fortnightly in 1967. Then it appeared as an evening daily for quite some time and finally converted into a full-fledged daily newspaper. It is published in Bengali, with online and English weekly editions, and is read in the city of Kolkata and the surrounding area. Although, a party mouthpiece, it covers wide range of issues including travel, literature, science, and technology. The principal catalyst behind transformation of *Ganashakti* was Saroj Mukherjee, a freedom fighter and CPI(M)'s state secretary during the 1980s. After Mukherjee's death his efforts were carried on by Anil Biswas, who was editor of the daily. Since 1 September 2004, Ganashakti has published an online version.

Goldnews actually started as telenews from 2003. In 2008 it became a full fledged online newspaper created by Goldnews (Barrackpore, Kolkata) which is under United Diginet.

Samayik Prasanga was started under the stewardship of Taimur Raja Chowdhury from 15th April, 1978 from Silchar as a fortnightly. From 11 May 2003 it is being published as a daily newspaper. Later it has also published the online version.

Sangbad Pratidin is a Bengali newspaper published from Kolkata. It started publishing from 9 August 1992, owned by Swapan Sadhan Basu. This was the first Bengali newspaper to start an online edition.

Century Sangabad was created by Century Computers in Kolkata in 1997. It is the online newspaper which does not have a print version.

Uttarbanga Sambad is a Bengali language broadsheet published from Siliguri. It is the largest circulating daily in North Bengal and has a share of about 80% out of all other dailies in the region. *Uttarbanga Sambad* was started on 19 May 1980 in a small letter press in Siliguri. Due to its huge popularity, in 1981 web offset press was installed. Computerised typesetting was introduced in the year 1985. Afterwards hi-speed 4-colour web offset press was introduced. *Uttarbanga Sambad* is now printed simultaneously from Siliguri, Cooch Behar and Malda. Since 2 April 2003, Uttarbanga Sambad has published an online version.

Presentation of News Items

Aajkaal: News items are separated in different sections. Some important news are published partially on the first page or home page. For getting full news readers have to click on that and a new window will come with full information.

Abasar: Different sorts of daily news along with some *special* news are published in separate sections. To view the news in full, readers have to click on the particular section of their choice or interest. Another very interesting feature is the link to famous Bengali Encyclopaedia 'Bharatkosh' along with search option.

Anandabazar Patrika : News items are separated in different sections. Some important news is published partially on the first page or home page. To access full news readers have to click on the link '*bistarita*' (expanded) which is given at the end of partial news.

Bartaman: News items have been presented in different sections. Some important news is given partially on the first page or home page. For accessing news in detail readers have to click on the link '*bishad*' (elaborateness) which is given at the end of partial news.

Daily Desher Katha : News items are arranged in different pages and sections. News presented on the first page providing partial information but for elaborate news one has to click on the hyperlinked page numbers that open to new windows.

Ganashakti: News items are separated by page numbers. For accessing detailed news readers have to click on the page numbers.

Goldnews: Different sections provide access to different news which again is hyperlinked under the heading of different sections.

Samayik Prasanga: It is the PDF version of print paper. News is arranged in pages as in print version.

Sangbad Pratidin : News items are separated in different sections. Some important news are provided partially on the first page or home page. One has to click on the particular news section to read it in detail on a new window.

Century Sangbad: Separate sections present different news items. Some headlines of burning topics are seen on home page. For detail information readers have to click on the headlines or news sections.

Uttarbanga Sambad: News items are separated in different pages and sections. Some important news is given partially on the first page or home page. To view full news readers have to click on the link '*bistarita*' (expanded) provided at the end of each partial news.

Option for Updated Breaking News

Only the *Anadabazar Patrika* provides the option for updated breaking news. But rest of the online Bengali newspapers have no such option .

Running Headlines

Aajkaal and *Bartaman* have running headlines at the upper side of the home page and *Anandabazar Patrika* has running headlines at the bottom section of home page. Rest of the online Bengali newspapers have no such feature.

News Sections

Aajkaal: The news sections are – First page, Kolkata, Bengal, India, Foreign, Editorial, Sports, Old edition and Classified arrangement.

Abasar : Different news sections found here are – First page, City information, Recreation, News, Law and Administration, Science and Technology, Art and Literature, Society and Culture, Health, Environment, Women and About *Abasar*.

Anandabazar Patrika : The news sections are – First page, Kolkata, State, Uttarbanga, Dakshinbanga, Bardhaman, Purulia, Murshidabad, Medinipur, National, Foreign, Business, Sports, Health, Environment, Editorial, Today Crossword, Comics, Feedback, Archives, About us, Advertisement rates and Font problem.

Bartaman : The news sections are – First page, Kolkata, Uttarbanga, Dakshinbanga, State, National, International, Sports, Editorial, Spiritual speech, Horoscope, Dollar-Pound, Rate of gold-platinum, Calendar, Share market rate, Archives, Public opinion, Communication, Advertisement, Font problem.

Daily Desher Katha: Available news sections are – Tripura, About us, Editorial, Agriculture, Science, National, International, Opinions.

Ganashakti: Here the page numbers have been mentioned. Generally the 5th page highlights the news of districts, 6th page highlights the news of Kolkata and adjacent areas and 7th page highlights the news of North Bengal.

Goldnews : The news sections are – First page, Editorial, Sports, Business, Entertainment, Health, Children's page, Foods, Horoscope, Share market, Fashion, Photo gallery, Video, Puzzle, Quiz, Jokes, Opinion, Communication, Classified, Advertisements rates.

Samayik Prasanga: No separate or particular news section mentioned except archive but page numbers have been mentioned.

Sangbad Pratidin : News sections are arranged in the following way : About us, National, International, Kolkata, Districts, Sports, Mahanagar, Editorial, Entertainments, Calendar and Contact us.

Century Sangabad: The news sections are - Miscellaneous scoop news, Sports, Songs, Cooking, Arts and Video, Horoscope, Health, Literature and Culture.

Uttarbanga Sambad : Available news sections are – About us, Comics, Advertisement, Editorial, Share market, Weekly horoscope, Archive, Emergency service, Your opinion, Bengali font. Besides this page numbers have been mentioned.

Arrangement of Pages

Page numbers are not mentioned except on the first pages of *Aajkaal*, *Abasar*, *Anandabazar Patrika*, *Bartaman* and *Gold news*.

Total page numbers are mentioned in *Daily Desher Katha*, *Ganashakti*, *Samayik Prasanga* and *Uttarbanga Sambad*.

There is no mention of page number in *Sangbad Pratidin* and *Century Sangbad*.

Hyperlink

All the 11 Bengali online newspapers have options for hyperlinks. Generally the full news may be accessed through hyperlinks which have been provided with the heading of different news sections or page numbers.

News Search Option

There is no such news search option present in the discussed online Bengali newspapers. Readers who want to read particular information have to click on the particular news section. But *Abasar* provides the link of 'Bharatkosh'— a Bengali encyclopaedia where the search option is available. Any reader who wants to search any topic has to click on the provided list of Bengali letters according to the first Bengali letter of his/her search term.

Provision for Back Issues

Provision for back issues or archives is available in *Aajkaal*, *Anandabazar Patrika*, *Bartaman*, *Samayik Prasanga*, *Sangbad Pratidin*, *Century Sangbad* and *Uttarbanga Sambad*. But there is no provision for back issues in *Abasar*, *Ganashakti*, *Gold news* and *Daily Desher Katha*.

RSS Feed

None of the discussed online Bengali newspapers has the option for RSS feed.

Focus Area of News

Abasar has a strong focus on women.. Whereas political news always gets highlighted in *Daily Desher Katha*. *Uttarbanga Sambad* has a focus on the various aspects of North Bengal. Rest of the online Bengali newspapers have no focus area as such.

Bengali Font

Software for reading Bengali font has been provided as a link in *Aajkaal*, *Abasar*, *Anandabazar Patrika*, *Bartaman* and *Uttarbanga Sangbad*. No link of software for reading Bengali font has been provided in *Daily Desher Katha*, *Ganashakti*, *Goldnews*, *Samayik Prasanga*, *Sangbad Pratidin* and *Century Sangbad*.

PDF Version

PDF version is available in *Aajkaal*, *Anandabazar Patrika*, *Bartaman*, *Daily Desher Katha*, *Ganashakti*, *Samayik Prasanga* and *Sangbad Pratidin*. But it is to be mentioned here that *Daily Desher Katha*, *Ganashakti* and *Samayik Prasanga* are actually the PDF version of print newspapers which have been uploaded on individual websites. On the other hand, *Abasar*, *Goldnews* and *Century Sangbad* are exclusively the online Bengali newspapers which have no print version.

Photography or Image

Ganashakti and *Sangbad Pratidin* are the only two newspapers that have

provision for enlarging images.

IMPACT OF ONLINE BENGALI NEWSPAPERS

A large number of net users in Bangladesh read Bengali newspapers of India online. Bengali NRIs are another beneficiary of online Bengali newspapers. They can keep themselves updated about political, social, cultural issues only by accessing the internet. Hence, they find it easier to access Bengali newspapers online to get the latest news. Moreover the online version or online Bengali newspapers give them the taste of home which they do not get from print or online English newspapers. (<https://drtc.isibang.ac.in/bitstream/handle/1849/108/Bf.pdf>)

There are also many internet savvy Bengalis who do not have the time to check printed versions. To them, the online version serves as a perfect channel to stay informed about current events. This has consequently enhanced the scope and readership.

The online Bengali newspapers are gradually gaining popularity. The printed paper risks losing revenue generators such as classified and personnel ads, and thus runs the risks of becoming less profitable. Competitors now fight in the same field for the users' attention and the limited advertising money available. The World Wide Web allows the existing and new media outlets compete for advertising revenues, and also provides a new range of advertising based services. Ads can be put on free communication services, on search engines and directory services, on free homepage services and e-commerce sites. Thus online Bengali newspapers have evolved into a profitable industry. Online versions are attracting loads of advertisements. The publishers of online versions or online Bengali newspapers are constantly trying to improve the quality and capture the Bengali news market.

(<http://home.medewerker.uva.nl/e.dewaal/bestanden/chapter4.pdf>)

ADVANTAGES

Online Bengali newspapers have some advantages over the print version. They are:

- Online newspapers are seamless. As a result there is no need to cut paragraphs or quotes. A story can be as long as it can be, and features, related articles, maps and charts can be linked to it. They can also add technologies such as full screen photos, videos, forums and polls that users can participate in.
- All the online Bengali newspapers are free of charge, often updated (e.g. *Anandabazar Patrika*) throughout the day, easily accessible to everyone with an Internet connection: and they can be visited while working at one's PC.
- Online editions lend themselves to use as a 'research' medium for updated or in-depth information for those who are motivated to process it.

- More stories on the front page are available in the online editions.
- Online editions also offer links to internal archives.
- The online configuration encourages users to control the flow of information by selecting stories of particular interest.

LIMITATIONS

- Online Bengali newspapers sometimes fail to open because of problem of accessing the Bengali font.
- Online edition of Bengali newspapers have so far not included the RSS feed option.
- The updated breaking news is not provided except by Anandabazar Patrika.
- In online edition clicking and scrolling may draw readers away from the other topics, whereas reading an article in a printed edition does not make the surrounding stories on that page or spread invisible.
- Bengalis are very fond of reading newspapers with their morning cup of tea but to read an online newspaper they have to turn a computer, dial on the internet provider, search for the newspaper they want and scroll and click through all the articles and photos. In the mean time the charm of reading newspaper with the morning cup of tea may be spoiled. In this regard, another problem which exists is that most of the online Bengali newspapers are not updated for the day before 10 a.m.

CONCLUSION

The advent of Internet has not spared the newspapers. The newspaper publishers all over the world are facing the challenge to compete with the electronic media. To survive in this competition they have started to publish the online versions of newspapers. Bengali newspapers in India are not far from this latest phenomenon. From the beginning of 21st century the online version of Bengali newspapers are available but few succeed in attracting traffic. Newspapers that have a sense of responsibility in circulating authentic pieces of information win the race. Good online Bengali newspapers should publish verifiable news in real-time i.e. live. This helps NRI Bengalis access news whenever they need. The online Bengali newspapers should also incorporate updates and RSS feeds to help people access news in real time. In this way the prospect of online Bengali newspapers would be increased. Already the online versions are attracting loads of advertisements. This would help the publishers to incorporate more features like RSS feed to compete with electronic media. Thus online Bengali newspapers would increase the global circulation.

REFERENCES

1. *Aajkaal*. www.aajkaal.net (accessed on 11.01.2010)
2. *Abasar*. www.abasar.net (accessed on 11.01.2010)

3. *Anandabazar Patrika*. www.anandabazar.com (accessed on 11.01.2010)
4. *Bartaman*. www.bartamanpatrika.com (accessed on 11.01.2010)
5. *Century Sangabad*. www.suprovat.com (accessed on 11.01.2010)
6. *Daily Desher Katha*. www.dailydesherkatha.com (accessed on 11.01.2010)
7. *Ganashakti*. www.ganashakti.com (accessed on 11.01.2010)
8. *Goldnews*. www.goldnewsindia.com (accessed on 11.01.2010)
9. Online newspaper. http://en.wikipedia.org/wiki/Online_newspaper (accessed on 17.01.2010)
10. *Samayik Prasanga*. <http://samayikprasanga.karimganj.com> (accessed on 11.01.2010)
11. *Sangbad Pratidin*. www.sangbadpratidin.net (accessed on 11.01.2010)
12. Schonbach, K, Waal, E D, Lauf, E. Online and print newspapers: their impact on the extent of the perceived agenda <http://home.medewerker.uva.nl/e.dewaal/bestanden/chapter4.pdf> (accessed on 17.01.2010)
13. Shankar, S S U. 2000. Online newspapers and magazines: aspects and issues <https://drtc.isibang.ac.in/bitstream/handle/1849/108/Bf.pdf> (accessed on 17.01.2010)
14. Sinha. 2007. Online newspaper readership in India. www.pluggd.in/india-digital-media/online-digital-newspaper-readership-in-india-14-use-it-as-a-tool-to-overcome-homesickness-574 (accessed on 16.01.2010)
15. Speech community. www.lisindia.net/Bengali?beng_Speech.html (accessed on 15.01.2010)
16. *Uttaranga Sambad*. www.uttarbangasambad.com (accessed on 11.01.2010)

ACCESS TO INDIAN ONLINE NEWSPAPERS: PROBLEMS AND PROSPECTS

Avijit Chakrabarti

ABSTRACT

An online newspaper, also known as a web newspaper, is a newspaper that exists on the World Wide Web or Internet, either separately or as an online version of a printed periodical. Going online created more opportunities for newspapers in presenting breaking news in a timelier manner. The credibility and strong brand recognition of well-established newspapers, and the close relationships they have with advertisers, are also seen by many in the newspaper industry as strengthening their chances of survival. The movement away from the printing process can also help decrease costs.

Indian newspapers have 230 years of history and the first newspaper the weekly 'Bengal Gazette', also known as Hicky's Gazette was published in January 1780 by James Augustus Hicky, an Englishman. India has more than 2000 newspapers in 21 languages with a combined circulation of 88 million, and some of leading dailies are The Times of India, The Hindustan Times, Malayala Manorama Dainik Jagran, The Telegraph, Asian Age, Indian Express, Economics Times, The Pioneer and many more.

Indian newspapers cater to the demand for information on local issues, politics, events, celebrations, people and business. Information about holidays, vacations, resorts, real estate and property together, finance, stock market and investments reports, theatre, movies, culture, entertainment, activities and events are all covered in Indian newspapers.

In India, major newspapers went online to provide latest and most updated news to the net savvy people. Some newspapers even provide e-Paper which is regarded as the digital replica of the newspaper.

The aim of this paper is to take a look on the newspapers published from different states of India and in different languages with problems and prospects with respect to access.

The demand for online newspapers has been increasing for the past few years with the growing reach of Internet. India's more than 70 million Internet users comprise approximately 6.0% of the country's population and about 4 million people in India have access to broadband Internet as of 2008 figures.

Online newspapers are much like hard-copy newspapers and have the same legal boundaries, such as laws regarding libel, privacy and copyright.

Keywords: Online newspapers, Indian newspapers, problems of online newspapers, prospect of online newspaper, future of online newspapers

INTRODUCTION

According to Napoleon Bonaparte (1769-1821), 'Four hostile newspapers are more to be feared than a thousand bayonets'. A newspaper is a publication containing news, information, and advertising. General-interest newspapers offer feature articles on political events, crime, business, art, entertainment, society and sports. Most traditional papers also feature an editorial page containing columns that express the personal opinions of writers. Supplementary sections mostly contain advertising, comics and coupons. A wide range of material has been published in newspapers, including editorial opinions, criticism, arguments, obituaries' amusement features such as crosswords, Sudoku and horoscopes; weather news and forecasts; advice, gossip, food and other columns; critical reviews of movies, plays and restaurants; classified advertisements; display advertisements, editorial cartoons and comic strips (Wikipedia). The new ICT (Information Communication Technology) came as both an opportunity and a risk from the point of view of the traditional printed newspapers. As a form of computer-aided communication, the WWW (World Wide Web) is equally a competition for the print media. Its technical potential greatly surpasses that of the printed newspapers in a number of ways. WWW has the advantages of being interactive, multimedia, of providing internal and external networks and offering selection functions, the possibility of regular updates, access to archives, rapid access to a large number of newspapers, and being paperless, thus creating no problems of waste disposal. One advantage left to newsprint is that reading it does not require any sophisticated, bulky technical equipment. This offers the reader a high level of flexibility: newsprint can basically be read in any place at any time. The reader can absorb the information offered at his own pace. Even the fact that the reader can touch and feel the printed paper while turning the pages may be of some importance. On the one hand, WWW presents a threat to the traditional distribution system; on the other hand it gives publishing houses the opportunity to offer up-to-date information, advertisements and additional services via a further communication channel (Van Oostendorp 1998). The aim of this paper is to take a look on the online newspapers published from different states of India and in different languages with its problems and prospects with respect to access.

SCOPE OF THE PAPER

The scope of the paper is restricted to a limited study of online newspapers published from India. India has 28 states with 22 official languages. The paper focuses on the current status of online newspapers in India with a brief history of newspapers in India. The paper does not provide the list of online news available on the sites of news channels. Paper also does not include the online newsmagazines published from India.

HISTORY OF NEWSPAPERS IN INDIA

Newspapers in India started with William Bolts, an ex-employee of the British East India Company who attempted to start the first newspaper in India in 1776. In 1780, James Augustus Hicky started *Bengal Gazette*, a two-sheet newspaper

that publicised the private lives of the 'sahibs' of the Company. In the newspaper, Hicky even dared to mount insulting attacks on the Governor-General and Chief Justice, for which he was sentenced to one year in prison and fined Rs.5, 000, which finally drove him to poverty. These were the first tentative steps of journalism in India. B. Messink Welby and Peter Reed were workable publishers of the *India Gazette*, unlike their infamous predecessors. The colonial establishment started the *Calcutta Gazette*. It was followed by another private initiative, the *Bengal Journal*. The *Madras Courier* was started in 1785 in the southern stronghold of Madras. Richard Johnson, its founder, was a government printer. Madras got its second newspaper when, in 1791, Hugh Boyd, who was the editor of the *Courier* quit and founded the *Hurkaru*. Tragically for the paper, it ceased publication when Boyd passed away within a year of its founding. It was only in 1795 that competitors to the *Courier* emerged with the founding of the *Madras Gazette* followed by the *India Herald*. The latter was an 'unauthorized' publication, which led to the deportation of its founder Humphreys. The *Madras Courier* was designated the purveyor of official information in the Presidency. In 1878, *The Hindu* was founded, and played a vital role in promoting the cause of Indian independence from the colonial yoke. Its founder, Kasturi Ranga Iyengar, was a lawyer, and his son, K. Srinivasan assumed editorship of this pioneering newspaper during the first half of the 20th century. Bombay surprisingly was a late starter. *The Bombay Herald* came into existence in 1789. Significantly, a year later a paper called *The Courier* started carrying advertisements in Gujarati. The first media merger of sorts happened with *The Bombay Gazette*, which was started in 1791, merging with *The Bombay Herald* the following year. Like, *The Madras Courier*, this new entity was recognized as the publication to carry 'official notifications and advertisements'. Compared with many other developing countries, the Indian press has flourished since independence and exercises a large degree of independence (Stephens 2008). Among the newspapers from Bengal, *Yugantar* and *Amritbazar Patrika* played a crucial role in freedom struggle. Today India has more than 2000 daily newspapers in 100 languages with a combined circulation of 88 million and some of leading dailies are *The Times of India*, *The Hindustan Times*, *The Telegraph*, *Asian Age*, *Indian Express*, *The Economic Times*, *The Pioneer* and many more. With the introduction of the Internet, Web-based 'newspapers' have also started online-only publications. In India, major newspapers went online to provide latest and most updated news from them. Some newspapers even provide E-Paper which is regarded as the digital replica of the newspaper. The demand for online newspapers has been increasing for the past few years with the growing reach of Internet. India's 70,000,000 Internet users comprise approximately 6.0% of the country's population and about 4,010,000 people in India have access to broadband Internet as of 2008 figures (Wikipedia).

ONLINE NEWSPAPERS

An online newspaper, also known as a web newspaper, is a newspaper that exists on the WWW or Internet, either separately or as an online version of a printed periodical. Going online created more opportunities for newspapers in presenting breaking news in a timely manner. The credibility and strong brand

recognition of well-established newspapers, and the close relationships they have with advertisers, are also seen by many in the newspaper industry as strengthening their chances of survival. The movement away from the printing process can also help decrease costs. Online newspapers are much like hard-copy newspapers and have the same legal boundaries, such as laws regarding libel, privacy and copyright, also apply to online publications in most countries, like in the UK (Wikipedia). Some newspapers have attempted to integrate the internet into every aspect of their operations, i.e., reporters writing stories for both print and online, and classified advertisements appearing in both media; others operate websites that are more distinct from the printed newspaper. Physically on the basis of publication policy, online newspapers are of three types –

- a. Online only newspapers – Online only paper is a paper that does not have any hard copy connections. It is independent web only newspaper.
- b. Hybrid newspapers – There are some newspapers which are predominantly an online newspaper, but also provide limited hard copy publishing.
- c. Soft-copy news sheet – A news sheet is a paper that is on one or two pages only. Soft-copy sheets are like online newspapers, in that they have to be predominantly news, not advert or gossip based. These sheets can be updated periodically or regularly, unlike a newspaper. They must also like a newspaper be regarded as a news outlet by media groups and governments.

(Wikipedia)

There are some advantages of online newspapers for their users: online newspapers are still mostly free of charge, often updated throughout the day, easily accessible for everyone with an Internet connection; and they can be visited while working at one's PC. As of 2009, the collapse of the traditional business model of print newspapers has led to various attempts to establish local, regional or national online-only newspapers - publications that do original reporting, rather than just commentary or summaries of reporting from other publications (Van Oostendorp 1998).

PRESENT STATUS OF ONLINE NEWSPAPERS IN INDIA

India has 28 states with 22 official languages. Online newspapers in all major language are published from India. Let's take a view of these newspapers with respect to different states.

Andhra Pradesh

Telegu – Eenadu, Sakshi, Vaarttha, Andhra Bhumi, Andhra Prabha, Andhra Jyothy, Leader, Netinizam, Rakshana, Pramukandhra, Kalanjali, Netibharatam, Prajasakti, Kaliyuga Narada, Praja Darbar, Viniyogadharula Vedika

Urdu – The Siyasath Daily, The Etamaad Daily, The Munsif Daily

Hindi – Praja Sahayata

English – E-Sameeksha, Takeed, Today Freedom, Tirupati Times, The Hindustan Times, Consumer Guide, Pathanjali News

Arunachal Pradesh

English – The Arunachal Times

Assam

Assamese – Dainik Agradoot, Dainik Jugasankha, Samayika Prasanga, Sentinel Assam

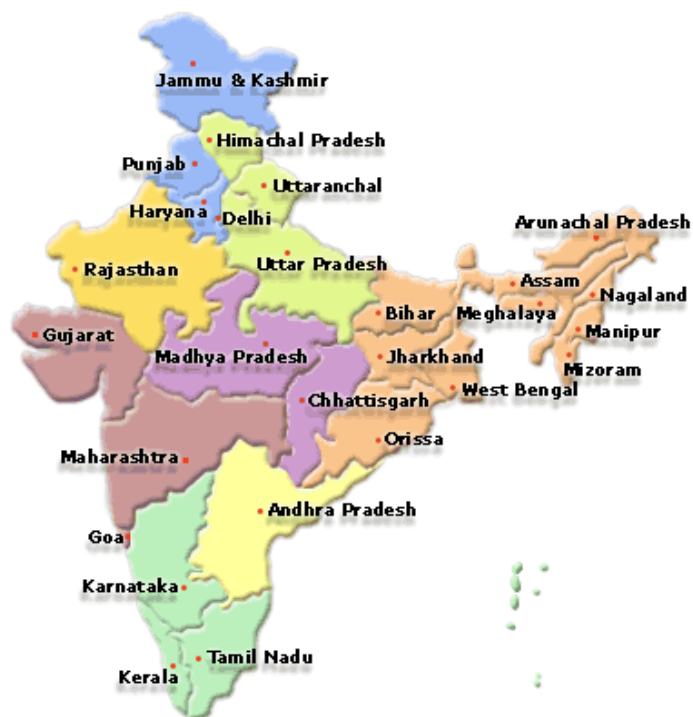
English – The Assam Tribune, The Sentinel, Voice of Assam

Bihar

English – The Bihar Times, Patna Daily

Chhattisgarh

Hindi – Dainik Bhaskar, News Guru India



Goa

English – Navhind Times, Gomantak Times

Gujrat

Gujrati – Divya Bhaskar, Jai Hind Daily, Naubat, Sandesh, Sambhaav

Hindi – Loktej

English – Lok Prabha

Haryana

Hindi – Sach Kahnoon

Himachal Pradesh

English – The Himetimes

Jammu & Kashmir

Urdu – Roshni, Etalaat, Mashriq Kashmir, Neab International, Khidmat, Wadi ki Awaz

English – Daily Excelsior, Greater Kashmir, Kashmir Observer, The Kashmir Times, The North Lines, J & K News

Jharkhand

English – Ranchi Express

Karnataka

Kannada – Janatha Vani, Kannada Prabha, Praja Pragathi, Praja Vani, Udayavani, Janatha Madhyama, Vishva Kannada, Prajodaya

English – Deccan Herald, Islamic Voice, Sahil online, Kranti Daily

Kerala

Malayalam – Deepika, Kerala Kaumudhi, Madyamam, Malayala Manorama, The Jasnews, Deshabhimani, Mathrubhumi, Kerala Online

English – Cochin 2 day, General Daily, New Kerala

Madhya Pradesh

Hindi – Nai Duniya

Maharashtra

Marathi – Pudhari, The Mumbai Samachar, Maharashtra Times

Urdu – Aurangabad Times, Inquilab

Hindi – Dainik Aikya

English – Indiatimes, Mumbai Mirror, DNA, Punecity, Raza Academy

Manipur

English – The Sangai Express

Meghalaya

Khasi – Mawphor

English – The Shillongtimes

Nagaland

English – Nagaland Post, The Morung Express

Orissa

Oriya – Dharitri, Orissa Sambad, Odisha, The Samaj

Hindi – Sanmarg

English – Kalinga Times

Punjab

Punjabi – Ajith Jalandhar, Sanju Savera, Parvasi

Rajasthan

Hindi – Pratahkal, Khaskhabar, Rajasthan Patrika, Myjhunjhunu

English – Khabar Express

Sikkim

English – Haalkhabar, Sikkim Express

Tamilnadu

Tamil – Daily Thanthi, Dina Karan, Thamil Sudur, Dinamalar, Akash Jyoti, That's Tamil

English – Patrika Online, South India Times, The Hindu

Tripura

Bengali – Daily Desher Katha

English – Tripura News, neindia

Uttaranchal

Hindi – Amar Ujala

Uttar Pradesh

Urdu – United News Network, Urdu net

Hindi – DLA, Rajdhani Times, Sahara Samay, UP Khabar, Deshbandhu

English – Agra News, Day after India, News Track, Public News

West Bengal

Bengali – Aajkaal, Uttarbanga Sambad, Bartaman Patrika, Ganashakti, Anandabazar Patrika, Sangbad Pratidin, Daily Desher Katha, Century Sambad, Gold News, Amder Malda, Abasor

English – Cal online, The Telegraph, The Statesman, Darjeeling Times

Delhi

Hindi – Navbharat Times

English - The Times of India, The Hindusthan Times

PROBLEMS OF ONLINE NEWSPAPERS

Today Indian online newspapers are facing many problems. Some of these are –

- a. The number one problem with online newspapers published from India is the font problem. Basically the problem arises in the case of Indic languages. Sometimes people need special software to read these newspapers. It is not possible for everyone to download software for each newspaper in a particular language. For example, one person can read Anandabazar Patrika by downloading Bitstream's Web Font Player software. But to read Bartaman Patrika, he/ she have to download another software though both are in Bengali language.
- b. The second problem is the presence of loud and stylistically unpredictable advertisements on the main page. Readers don't visit these sites to be overwhelmed by hectic colors and dizzying arrays of advertisements. Rather, they visit to read news and be informed. These ads distract and they destroy the design and the layout of the pages. The worst offenders are animated advertisements. Trying to read the headlines with one of these animated advertisements next to it is like trying to read the headlines of a real newspaper within a fish market.
- c. In almost every case the layout and design of online newspapers is very poor. Very few online newspapers employ intelligent use of grid, color and content hierarchy. Most have a flat aspect with no visual clues as to how the page and its information are supposed to be consumed.
- d. With most online newspapers, the site navigation is horribly confusing and inconsistent. Furthermore, the link text on the page is often disguised, easily confused with non-link text, and provides little or no visual feedback when used.
- e. Online newspapers often have questionable typography and paragraph configuration. Some have quite a horrible mismatch of font styles, coupled with poor font selection. Proper typography is more important in the low-resolution online environment than it is in the high-resolution environment of print. Too often, online newspapers have bad line-height (leading) in their paragraph copy, making it more difficult to read on the monitor.
- f. Although the Internet can help online publishers save the cost of printing and delivering the hardcopy newspaper, online newspapers have yet to generate sizeable revenues by charging their readers.
- g. In India, ICT infrastructure is very poor. To access online newspaper the first thing we need is a computer with internet connection. It has been found that most of the villages in our country do not have any internet connection. Electrical power supply is also erratic.

PROSPECTS OF ONLINE NEWSPAPERS

Online newspapers in India have much business potential because -

- a. The cost of starting an online newspaper is much lower than that of starting a traditional print newspaper. For newspapers already available

in a print version, the added cost for maintaining an online version is really minimal. Considering the limitless space provided by the Internet and the increasing cost of newsprint, online publishing is much more cost effective.

- b. The country's population today has crossed 1.1 billion; out of them almost 70% live in rural areas. The literacy rate which was around 17% at the dawn of independence, stood as 66% as per census of 2001 and would have further improved since. In absolute numbers, people are gradually rising above the poverty line and with passage of time, more literate and educated people and economic growth are reasonably expected. This change and progress will improve capacity to purchase and read. Though the computer literacy rate is still very low and only 6.12 % of total population are computer educated. More people want to gain computer literacy with time (Lalwani 2009). Probably in near future more people will get accustomed to live and work in a networked environment.
- c. Government of India now gives emphasis on the development of infrastructure including ICT. This will create facilitate spread of the online newspapers.
- d. An online newspaper has some advantages which include:
 - given the short concentration span of Internet users, concise text captures more attention;
 - forums create an environment where people entertain people, making the media a sideshow;
 - feelings expressed in the forum cannot be matched by journalistic interpretation and writing skills;
 - the Internet users include an untapped generation of non-readers who will try everything, only once, which makes scouting for them out not worth the effort.

These advantages plus others will no doubt give online newspapers enough reasons to stay and be further explored. Indian online newspapers can adopt these advantages to increase their readership base.

FUTURE

People believe that the printed newspapers would not be replaced by the electronic newspapers. They believe that electronic newspapers would be supplementary to the printed copies (Bogart 1984). But the publishers believe that both print and online newspapers would flourish together (Mueller 1995). In general, Indian online newspapers seem to be more optimistic about their future. The current development of Web newspapers is underlined by newspaper publishers' desire to turn the Internet into an opportunity to help reverse the trend in declining readership rather than a threat to the print newspaper. The newspaper industry seems to be more confident about the future with the development of online publishing (Boczkowski 2002). Publishers and online editors believe there is little chance for the print newspapers to be replaced by

the electronic version. Instead, online publishing is seen as having the potential to open up broader prospects for the newspaper industry (Fransen 1998). The future of a new medium depends on whether it is simply a replica of an existing one and to what extent it can add value to it. In this regard, online newspapers clearly enjoy some advantages over the print newspapers. First and foremost, online newspapers are gradually building up a readership of their own. The readership of online newspapers is composed of a special group of newspaper readers who differ from those who read the same newspaper in its hard copy. They are more likely to be those who read more than their local daily and need to read other newspapers but do not have direct access to hard copies. Whatever the reasons, online newspaper readers are more likely to help swell the ranks of existing readers rather than reduce them. Also, online newspapers provide new services that could not be a part of traditional newspapers (D'Haenens 2004). The searchable archive and classified ads can help each newspaper become an information databank in addition to its role as a deliverer of news. The hyperlinks have changed the newspaper from a single source of information into a hub of information networks without a clear ending point.

CONCLUSION

New technologies can bring new opportunities as well as threats to existing media. The newspapers play an important role in crystallizing public opinion. The forums, chat facilities and e-mail contacts allow greater interactivity among the community members and between the readers and editors than the traditional print newspapers (Bogart 1985). As far as business is concerned, very few online papers are actually making money. Compared with the print paper, the online paper is clearly at a disadvantage in maintaining the traditional sources of newspaper revenues - advertising and subscription charges. Newspaper advertisers are mainly interested in local customers within their reach, but online readership tends to be away from the local market. Only multinational corporations and companies catering to the international market may find the online paper useful to carry their advertisements. As a result, small local online newspapers find it hard to generate advertising revenues. It may be still too early to conclude that advertising will not become a major source of revenue for online papers. On the one hand, it takes time for advertisers to assess the online readership and the effectiveness of online advertising. On the other hand, online newspapers have to make full use of the advantages provided by the Internet to offer better services for advertisers. For instance, the limitless space provided by the Internet will allow advertisers to have their ads carried by online papers for as long as necessary. As a result, it may be more cost effective to advertise in online papers than in the print. In addition, online search engines will also make it more convenient for readers to select what they need. As far as subscription is concerned, the situation is not optimistic for online papers. The Internet culture is characterized by free information. Some newspapers are limited to certain uses such as access to archives. There should not be fixed means for online papers to generate revenues but papers with a large readership base and more specialized orientation tend to be more successful with advertising and subscription charges while smaller papers tend to rely on Internet services for revenues. Considering the relatively small amount of costs for online papers and

the additional benefits brought about by online papers such as promoting the print product, online papers should not be considered simply as money losing operations by the publishers. If online publishers can focus on turning out a value-added product rather than a replica of their print product, online papers will become cost effective operations with the development of new technologies and growth of the Internet community. That probably explains why few online papers have closed shop since they were started.

REFERENCES

1. Boczkowski, P. J. 2002. *The Development and Use of Online Newspapers: What Research Tells Us and What We Might Want to Know*. In *The Handbook of New Media*, pp.270-286, edited by L. Lievrouw and S. Livingstone (eds.) London: Sage
2. Bogart, L. 1984. *The Public's Use and Perception of Newspaper*. *Public Opinion Quarterly*: 709-719.
3. Bogart, L. 1985. *How US Newspaper Content Is Changing*. *Journal of Communication*: 82-90
4. D'Haenens, L., N. Jankowski & A. Heuvelman. 2004. *News in Online and Print Newspapers: Differences in Reader Consumption and Recall*. *New Media and Society*, vol. 6 (no. 3): 363-82
5. Franssen, M. 1998. *Electronic newspaper case study: Flexible user interface design for easier web site maintenance*. *Trends in Communication. Special Issue on Interface Technology: Enhancing the Quality of Life*, vol. 2 (no. 1).
6. Mueller, J. & D. Kamerer. 1995. *Reader Preference for Electronic Newspapers*. *Newspaper Research Journal*, vol. 16 (no. 3): 2-13
7. Lalwani, S. 2009. Computer literacy – India's key to progress and development. www.groundreport.com/Opinion/Computer-Literacy-Indias-key-to-Progress-Development-1/2890553 (accessed on 31.1.2010)
8. Stephens, M. 2008. *History of Newspapers*. [www.nyu.edu/classes/stephens/Collier's page.htm](http://www.nyu.edu/classes/stephens/Collier's_page.htm) (accessed on 31.1.2010)
9. Van Oostendorp, H. & C. Van Nimwegen. 1998. *Locating Information in an Online Newspaper*. *Journal of Computer-Mediated Communication*, vol. 4 (no. 1)
10. Wikipedia. Newspaper (html). <http://en.wikipedia.org/wiki/Newspaper> (accessed on 17.12.2009)
11. Wikipedia. Online Newspaper (html). http://en.wikipedia.org/wiki/Online_newspaper (accessed on 17.12.2009)

ACCESS TO ONLINE NEWSPAPERS OF INDIA: AN OVERVIEW

Dr. O.N. Chaubey

and

Dr. (Mrs.) R.Chandra

ABSTRACT

The number and use of online newspapers has increased tremendously. One of the reasons for the increase in the growth of online newspapers in India was the increasing literacy, as measured by the ability to read and understand any language. The literacy rate was estimated to have risen from 62.5 per cent to 70.6 per cent in the three years from 2002 to 2005. Interestingly, the increase was higher in rural (from 55.6 per cent to 64.6 per cent) than in urban (from 79.3 to 84.5 per cent) India. Readership growth was faster than literacy growth. While print media is going strong, it is interesting to see is the rise of digital /online newspaper consumption in smaller cities of India and also online versions of Indian newspapers are read in 62 countries. The reach of the readership of newspaper was higher in the southern, western, and northeastern States, which have higher literacy rates. This paper attempts to analyze the distribution pattern of online newspapers of India and also gives in detail a panoramic view of surveys conducted by various agencies on the influence of online newspapers in India.

INTRODUCTION

Indian newspapers have more than 400 years of history. James Augustus Hickey is considered as the "father of Indian press" as he started the first Indian newspaper from Calcutta, the *Calcutta General Advertise* or the *Bengal Gazette* in January, 1780. In 1789, the first newspaper from Bombay, the *Bombay Herald* appeared, followed by the *Bombay Courier* next year (this newspaper was later amalgamated with the Times of India in 1861).

The first newspaper in an Indian language was the *Samachar Darpan* in Bengali. The first issue of this daily was published from the Serampore Mission Press on May 23, 1818. In the same year, Ganga Kishore Bhattacharya started publishing another newspaper in Bengali, the *Bengal Gazetti*. On July 1, 1822 the first Gujarati newspaper the *Bombay Samachar* was published from Bombay, which is still extant. The first Hindi newspaper, the *Samachar Sudha Varshan* began in 1854. Since then, the prominent Indian languages in which papers have grown over the years are Hindi, Marathi, Malayalam, Kannada, Tamil, Telugu, Urdu and Bengali. India has more than 2000 daily newspapers in 100 languages with a combined circulation of 88 million.

The print media has responded to the new changes and challenges with its modernization. They have accepted the information technology, which resulted in better coverage with great speed and affordable price. The readership of newspapers is growing. The statistics also shows that the people prefer their regional language newspapers and that is why the regional newspapers are venturing out to bring editions from other cities where there is sizeable population of the respective language.

The Indian language papers have taken over the English press as per the latest National Readership Survey (NRS) survey of newspapers. The main reasons being the marketing strategy followed by the regional papers, beginning with *Eenadu*, a telegu daily started by Ramoji Rao. The second reason being the growing literacy rate. Increase in the literacy rate has direct positive effect on the rise of circulation of the regional papers.

The people are first educated in their mother tongue as per their state in which they live for e.g. students in Maharashtra are compulsory taught Marathi language and hence they are educated in their state language and the first thing a literate person does is read newspapers and gain knowledge and hence higher the literacy rate in a state, the sales of the dominating regional paper in that state rises.

ONLINE NEWSPAPER

Online newspapers were considered a revolution in mass media in the 1990s. Their characteristics of hypertext, interactivity, and multimedia necessitated a new model of journalism and reading behavior. Since the late 90s growth in development and readership has been dramatic with most large and medium newspapers now having an online presence. An online newspaper, also known as a web newspaper, is a newspaper that exists on the World Wide Web or Internet, either separately or as an online version of a printed periodical. Online newspapers are much like hard-copy newspapers and have the same legal boundaries, such as laws regarding libel, privacy and copyright.

Going online created more opportunities for newspapers, such as competing with broadcast journalism in presenting breaking news in a more timely manner. The credibility and strong brand recognition of well-established newspapers, and the close relationships they have with advertisers, are also seen by many in the newspaper industry as strengthening their chances of survival.

ONLINE NEWSPAPERS IN INDIA

The Internet is spreading across the country, and newspaper websites too seem to be making hay off it. The Web editions of Indian newspapers are now read even in smaller towns, and not just in the metros of the country. The demand for online newspapers has been increasing from the past few years with the growing reach of Internet. India's 60,000,000 Internet users comprise approximately 6.0% of the country's population and about 4,010,000 people in India have access to broadband Internet as of 2008 figures. ePapers provide the latest breaking news with a minute update.

The reach of online newspapers is practically endless and is not bounded by area or boundaries as not only one can get local but international exposure. In real terms, the landscape of online outreach has drastically changed. Another great reason for having an online presence is that, nobody reads yesterday's newspaper. However, a piece of online communication stays, typically, anywhere more than a day to years together. Probably, online presence stays till the storage runs out.

ANALYSIS OF ONLINE NEWSPAPERS OF INDIA

Methodology Adopted

This study has attempted to analyze the availability of online newspapers of India and their distribution pattern. There are innumerable sites on the Internet to show the availability of online newspapers in India. Since there is a lot of variation from one site to another, this study could identify only 294 online newspapers available and the same has been taken for analysis.

Statewise Analysis:

A notable feature is that Delhi has achieved the distinction of publishing 44 online newspapers and Tamilnadu comes next with 40 newspapers, Maharashtra with 34, followed by Karnataka having 22 followed by West Bengal , having 21 newspapers.

States	No. Online Newspapers	States	No. Online Newspapers
Andhra Pradesh	18	Maharashtra	34
Assam	11	Manipur	5
Bihar	4	Meghalaya	1
Chandigarh	1	Mizo	1
Delhi	44	Orissa	9
Goa	5	Punjab	7
Gujarat	12	Rajasthan	11
Haryana	3	Sikkim	2
Himachal Pradesh	1	Tamil Nadu	40
Jammu & Kashmir	9	Tripura	3
Karnataka	22	Uttar Pradesh	10
Kerala	14	West Bengal	21
Madhya Pradesh	6		

Language wise – Analysis:

For the convenience of people, the Constitution of India has recognized 22 languages as official languages of India. These are known as Scheduled Languages and constitute the Major languages of India. The language wise distribution of online newspapers is shown below: The language English scored the highest of having 114 online newspapers, followed by Hindi, the most widely spoken of India's official languages having 45 followed by Tamil with 25

newspapers available on online. The other official and other languages of India having their respective number of online newspapers are shown in the table.

LANGUAGES	NO. OF ONLINE NEWSPAPERES
Arabic	1
Assamese	8
Bengali	15
English	114
Gujarati	13
Hindi	45
Kannada	11
Malayalam	10
Marathi	12
Nepali	2
Oriya	8
Sanskrit	1
Punjabi	6
Tamil	25
Telugu	10
Tibetan	1
Urdu	11

Globally Popular Online Newspaper of India:

The Times of India (TOI) is a popular English-language broadsheet daily newspaper in India. It has the largest circulation among all English-language daily newspapers in the world, across all formats (broadsheet, compact, Berliner and online). It is owned and managed by Bennett, Coleman & Co. Ltd. which is owned by the Sahu Jain family.

In 2008, the newspaper reported that (with a circulation of over 3.14 million) it was certified by the Audit Bureau of Circulations as the world's largest selling English-language daily newspaper, placing as the 8th largest selling newspaper in any language in the world. According to the Indian Readership Survey (IRS) 2008, the *Times of India* is the most widely read English newspaper in India with a readership of 13.3 million. This ranks the *Times of India* as the top English newspaper in India by readership. According to ComScore, TOI Online is the world's most-visited newspaper website with 159 million page views in May 2009, ahead of the *New York Times*, *The Sun*, *Washington Post*, *Daily Mail* and *USA Today* websites.

Online Newspaper in Sanskrit:

Sudharma, the only Sanskrit daily newspaper in India, began its online version in June 2008. *Sudharma* daily sells around 3000 copies through subscription. The internet version of *Sudharma* is aimed at reaching a wider audience. The only income generated by *Sudharma* is from the 250 rupees of annual subscription. By going online *Sudharma* is planning to attract advertisers, which is essential for the survival of the only Sanskrit daily.

Surveys Conducted:

Readership:

By UGC:

A recent study has found out that the readership of online newspapers has also spread out to as many as 62 countries, in addition to India. The growth of online newspaper readership within India has been possible due to penetration of computers and easier availability of Internet connectivity even in small towns and cities, according to the study, which was carried out by the Department of Communication and Journalism, University of Pune and the study was funded by the University Grants Commission (UGC). Some of the findings of the study are as follows:

- The female English Internet newspaper readers: 13% (This has doubled from the previous study);
- The majority readership is in the age group of 21-40 years, and it is more in states like Maharashtra, Karnataka, Delhi, Tamil Nadu and Andhra Pradesh;
- Majority of the readers are highly educated i.e. 60%
- Browsing online newspapers mainly for news – 95%
- Browsing for Book/film/theatre reviews - 24.3%
- Advertisements - 9.10%.
- Reading more about business & economy – 69%, politics – 60%, and science & technology -60%.
- Sports news is favoured only by 37% of the users.
- Matrimonial advertisements by about 2.5% readers
- Readers of Cartoons - only 0.44%

Indian Readership Survey:

According to Indian Readership Survey – 2009, the Dainik Jagran is the most-read, local-language (Hindi) newspaper, with 55.7 million readers.

The National Readership Study:

The National Readership Study 2005 (NRS 2005) in India is the largest survey of its kind in the world, with a sample size of over 2,61,212 house-to-house interviews to track the media exposure and changing consumer trends in both urban and rural India - and of course the estimated readership of publications.

Highlights:

- The number of readers in rural India is now roughly equal to that in urban India.
- The time spent reading has gone up quite significantly though - from 30 minutes daily on an average to 39 minutes per day over the last three years. The increase has been sharp both in urban India (from 32 to 42 minutes daily) and in rural India (from 27 to 35 minutes daily).

ONLINE NEWSPAPER ARCHIVES:

Old newspaper articles are great resources and the following online newspapers are providing archives:

- The Times of India (1996-)
- The Hindu (2000-)
- Hindustan Times (1996-)
- The Indian Express (1997-)
- The Telegraph (1999-)
- The Statesman (1999-)
- The New Indian Express (2000 -)
- DNA (2005 -)
- Economic Times (1996 -)
- PunjabNewslines (2001 -)

CONCLUSION

Online newspapers are an important source of news for many people. The issue of relative importance to users of factors in content, interface design, and multimedia presentation is important. This study provided some insights for online newspapers and however, further research is recommended.

REFERENCES

1. Thiel, S. (1998). The online newspaper: a postmodern medium. *The Journal of Electronic Publishing*, 4(1). Retrieved 30 August, 2003 from <http://www.press.umich.edu/jep/04-01/thiel.html>
2. Herre van Oostendorp and Christof van Nimwegen (September 1998). "Locating Information in an Online Newspaper". *Journal of Computer-Mediated Communication* 4 (1). <http://jcmc.indiana.edu/vol4/issue1/oostendorp.html>.
3. Massey, B., & Levy, M.R. (1999). Interactivity, online journalism, and English - language Web newspapers in Asia. *Journalism and Mass Communication Quarterly*, 76(1), 138-151 McAdams, Melinda (July 1995).
4. "Inventing an Online Newspaper" *Interpersonal Computing and Technology* 3 (3): 64-90. http://www.eric.ed.gov/sitemap/html_0900000b8001fe35.html.
5. <http://oldcontent.newswatch.in/lead-story/6874.html>
6. "TOI Online is world's No.1 newspaper website". Times of India. July 12, 2009. <http://timesofindia.indiatimes.com/NEWS-India-TOI-Online-is-worlds-No1-newspaper-website/articleshow/4769920.cms>. Retrieved 2007-10-16.

7. "Dailies add 12.6 million readers: NRS". *The Hindu*. August 29, 2006. <http://www.hinduonnet.com/thehindu/holnus/001200608291820.htm>. Retrieved 2007-10-16.
8. "List of newspapers in the world by circulation". International Federation of Audit Bureaux of Circulations. June 30, 2008. http://en.wikipedia.org/wiki/List_of_newspapers_in_the_World_by_circulation. Retrieved 2007-10-16.
9. *The Times of India*. June 30, 2008. <http://www.newswatch.in/newsblog/1821>. Retrieved 2007-10-16.
10. *Frontline*, July 19, 2002. Dailies add 12.6 million readers: NRS. NRS Chennai, 2009 August 29. (Web.Archive.org).

GOOGLING NEWS AND VIEWS ONLINE

Dr. D.R. Gupta

ABSTRACT

On-line newspapers have emerged as a new source for the current information in the last few years. They bring to the reader's desktop the global information by overcoming all geographical limits and make them aware of all current news. Googling news and views over the Internet by many user groups including academic fraternity, governmental and commercial institutions and general public is the new trend worldwide. This paper discusses the various aspects and issues regarding online newspapers particularly with reference to India.

INTRODUCTION

The primary task of every library has always been to provide the reader with a source of information. Until recently it was mostly in the form of books, magazines and newspapers. Nowadays, the most important things for library users are: quick service, full collection, quick information on the required material as well as getting access to world information resources. The Internet now offers the capability to find information on everything that man has created and put on the network, and it offers unlimited access to a huge amount of information resources.

Information Technology is the backbone of today's information society. The advent of digital technology has effectively changed the nature of global communications, amplifying the interaction between populations and massively expanding the information load. With the introduction of World Wide Web many a newspapers and news magazines have started to publish their editions online. This helps a reader to access the newspapers and gather views of the world's public. This trend, in real sense, has helped develop an 'Information Democracy' among the users world over.

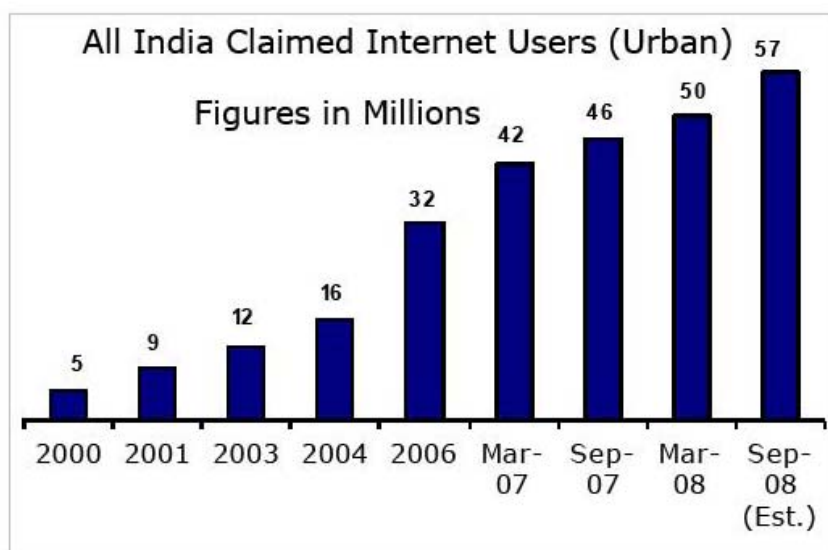
GROWTH OF ONLINE NEWSPAPERS

Online newspapers first appeared in the 1940's itself, albeit in a different form. Some radio stations in US sent newspapers using facsimile at home. They also made efforts to deliver them over phone lines to the user's home TV sets. Companies like Compuserv offered electronic editions of national newspapers in the 1980's. They also posted their newspapers on Bulletin Board Services. All these attempts were only on experimental basis and therefore, these attempts could not gain popularity as the cost of this service was too high and also there were only a few subscribers. Technologists then realised that online editions of the print media could not become a successful medium to deliver the expected results. Advancement in hardware and software technology and the advent of Internet during mid 1990's was a boon to the print media. Since then they started

to publish their online editions. As the World Wide Web became popular, new breeds of information providers, such as AT&T and Microsoft's Slate, started providing online news. Newspapers had to keep pace with this new development, the Internet.

According to American Journalism Review News Link, the number of online newspapers increased from very few to 1,300 in 1996. In June 1997, there were 3,622 online newspapers worldwide. In 1995, many of the newspapers in Europe and US began to publish their online editions. Editors of online medium considered that many political, social and sport events could be published online quickly, competing with TV and radio. They published more contents, including an index, a package of articles and photos that could rival broadcaster's work. Scientific journals were the first to get published online in 1995. Then the general interest magazines followed their footsteps.

When considering the trends in newspaper websites in 2000 and 2008, it is evident from the following graph that the Internet users have increased rapidly in India.



Source: I cube, 2008, IMRB Syndication

Figure 1: Steady rise in the numbers of Internet users in India

In, 2009, India has been ranked fourth among the top 10 nations in the world, in terms of Internet usage, with 81 million Internet users. United States leads the chart with 220 million Internet users followed by China (210 million) and Japan (88.1 million).

With the increasing quality, Broadband and Wi-Fi services on very nominal prices or sometimes even absolutely free, have been provided by various service providers. The number of internet users worldwide is expected to touch 2.2 billion by 2013 and India is projected to have the third largest online population during the same time, says a report. "India will have the third largest internet

user base by 2013 - with China and the US taking the first two spots, respectively," technology and market research firm Forrester Research said in a report.

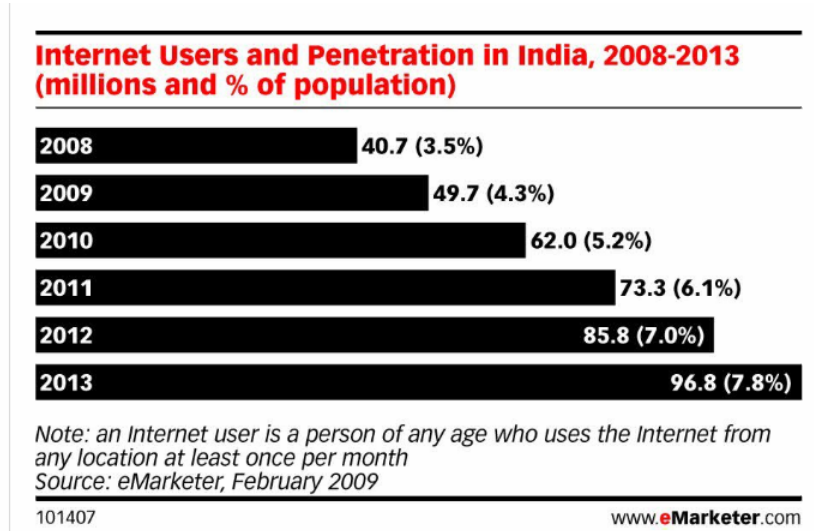


Figure 2: Internet users and penetration projections for India

What is surprising is the revelation that 90% of Internet users come from top 10 cities in India. Further, imagine that the 63% Internet users are not from Mumbai, Bangalore, Delhi, Calcutta, Chennai, Pune, Hyderabad, Ahmedabad, Surat and Nagpur. Only 41% who prefer to read English is another surprise.

The use of Mobile Internet is rapidly spreading, especially among the young generation. Mobile Internet is the system of interconnected network which is accessed through mobile devices such as a cell phone, PDA or other portable gadgets or browsing a mobile network such as Wi-Fi. The point is, access to Internet doesn't require desktop or fixed landline connection, Mobile Internet is also referred as 'Internet on Move' in some jargons. The latest Internet & Mobile Association of India (IAMAI) report on Mobile Internet usage in India shows that India had 57 million Internet users as on Sep-2008; the numbers are likely to grow rapidly.

ONLINE NEWS IN THE PRESENT INFORMATION AGE

Information dissemination has got a new dimension in this digital era. The transformation of the communications sector indicates a shift from mass broadcast media to interactive media. Printed matter is getting replaced by digitized information. Internet is providing its users more choice and control of content, and enables individuals to produce and distribute information. As a result, Internet news is proliferating; produced by individuals, freelance experts, advertising agencies and traditional news organizations. Most of the information can be accessed freely.

The World Wide Web's role in the aggregation and dissemination of information and news affects the printed media in various ways. The revenue base for the printed media gets threatened by decrease in the number of readership and advertising revenues, but at the same time many new opportunities for news production and distribution are also present. The transformation of newspaper publishing from a separate and independent enterprise into one of the products of multi-media production houses is a very significant phenomenon. The software and telecommunications industries, and also the television industry, compete and co-operate with newspapers on different levels, bringing new issues about access and freedom of information. These processes are taking place at a rapid speed that there is little chance for public intervention or corporate analysis and strategy building. The transfer of digitized information from the source to the destination is easy and less complicated when compared to the printed matter. With the help of the Internet technology, the information transfer has become so fast. The effort and time needed for the transportation to get the information has minimized to a great extent in this digital era. The user population of Internet has grown considerably; the population must be equivalent to that of the traditional media, if not more than that. All these factors have contributed to the emergence of the online newspapers and other material.

PLANNING AND ARCHITECTURE OF THE ONLINE MEDIA

There have been newspapers published for almost half a millennium now. Much effort has been put into developing a proper functional architecture for the digital newspaper library, as a part of an overall digital architecture. This will include all kinds of materials and the production of the digital versions as the digital original, the digital master and the digital image on the web and the textual versions for full text searching.

The Internet today is the world wide web of computers using a common language to publish material to computers with different architectures throughout the world. Hundreds of traditional media outlets - newspapers, magazines, and radio and television stations – are now producing Web versions of their publications. Many of them highlight the key words in blue colour and provide links to the matter on the same subject. Some online newspapers customize the news to fit each user's needs. Information is provided here along with photographs, audio and video clips thus providing a comprehensive information and news to the users. According to the online experts, if online newspapers are to reach the users of all types, then they have to change their mission, content, design and presentation of information. Many online newspaper or magazine websites also provide archival collection of their back issues.

Newspapers and other news material have been a passive medium in the past years. But after the publication of their online editions, users want responses from reporters, editors, online staff and departments and they want appropriate response. E-mail has become a great way to get the response and feedback. Online news gives its readers an alternate route to respond to particular columns or information. If a popular columnist or reporter receives many mails, an assistant can screen it and pass on the interesting messages or encourage the

columnist or the reporter to take part in an online discussion where they can post a response to a user's message and everyone can read it.

Various forums, chat rooms, Facebook and Buzz options also help the journalists to get the feedback and queries from their users. Since there are millions of pages of information on the Net, online newspapers are collecting, filtering, recording and distributing information at many levels and many places around the world.

Since there has been much unauthentic information provided in the Web – journalists have to prove the authenticity of their information. To improve the content of the online news material, they have to provide information that users can't get elsewhere, such as that on current happenings, or on a particular industry, sports etc. For a new publisher it is better to publish a popular section or a new supplement online, rather publishing the entire newspaper or a magazine. Depending upon the response, the number of hits and the feedback, a full-fledged online newspaper or a magazine can be published.

Online newspapers are obligatory resources. They act as a community encyclopedia. The online newspapers have to compete with other traditional as well as other online newspapers, and also with large news organizations, such as BBC, CNN etc. A few online newspapers have started updating information on a 24x7 basis.

FINANCIAL ASPECTS

Financing is still one of the main problems of online newspapers. Though they are not as capital intensive as their printed counterparts, even so they too require large outlays on a continuing basis (especially for publishers who as providers also offer readers access to Internet). Advertisers are still very skeptical and users unwilling to pay, as the project's user survey showed.

Online newspapers and other relevant material in today's digital environment are expected to provide various services that appear to be as effective as radio and television services. Online editions of newspapers have become more successful in the Western countries. In India it is slowly gaining popularity. The literacy level which was, until now, limited to young generations, have now increased in the older group of people. The World Wide Web allows the existing and new media outlets compete for advertising revenues, and also provides a new range of advertising based services. Advertisements can now be put on free communication services, on search engines and directory services, on free homepage services and on e-commerce sites. The difficulty in attracting users to a particular website results in growing advertising budgets for these online editions. Many e-commerce companies spent on an average over 8.6 million dollars per year each on advertisements, marketing and branding, of which a large part is going to existing print and broadcast media. The fragmentation of the online public gives more prominence to the gathered audiences of print and broadcasting.

News services on the Web come loaded with interactive features, guides to information sources and now many online newspapers are trying to build online

communities. Economically, the option of multiple revenue streams means that the entire publishing industry should turn to a database model, by which generated information or content can be marketed and sold in different formats. This offers interesting new opportunities for the printed media, which has many advantages in terms of a high trust quotient of its content, a well-organized news production model, and access to both readers and advertisers, but it also means that the Press needs to invest substantially in research and development. Long-term strategies and large financial commitments are needed, which may be hard to come by for small individual newspapers.

Internet allows user to interact on different subjects of their choices. This is an excellent advantage the online editions have over their print counterparts. There being low barriers to entry, they offer publishers the opportunity to develop additional revenue streams based on their core product, the collection and analysis of information. The interactivity of the medium has proven to be attractive for many, drawing the audience away from television to return to a largely text-based medium.

ADVANTAGES AND DISADVANTAGES

Advantages

The fact that several newspapers are available for reading on the Internet is very important for the masses. The advantages resulting from hypertext, automated searching and the availability of back numbers are also rated highly. The relative importance of feedback is always in line with the low utilization of internet usage.

An online publication allows the publishers to get the information to the reader much more quickly than the print newspaper. Since the online editions of newspapers are seamless, there is no need to cut paragraphs or quotes. A text can be as long as it can be, and features, related articles, maps and charts can be linked to it. They can also add full screen photos, videos, forums and polls that users can participate in and deliver their views and vote accordingly. They also allow for discussion beyond the Letter to the Editor page, for example, contact between journalists, editors and readers, and for relevant information regarding readers' preferences and choices.

There is possibility to update and upgrade the news or having even multiple hyperlinks which are linked with full screen photography and video. The sites of online newspapers also allow the users to browse and select news and advertisements at their own time and place. Some online newspapers attract users to allow them to create a series of web sites themselves.

Disadvantages

One serious disadvantage of online newspapers is that they usually provide only a selection of the articles available in the print newspaper. Only three of the online publications that participated in a survey included all articles from the print newspaper every day. Other disadvantages cited by many users are that online newspapers can not provide the same experience of reading as a print

newspaper. The download times are long and the "paper" lacks portability. There seem to be few technical problems connected with reading online newspapers and the cost of Internet access is not seen as a serious drawback.

It is hard to recognize online newspapers on the World Wide Web. Also the customizing of news to a user's preferences and searching for something on the online newspaper or magazine's site could harm the privacy of the users. This is because of the use of cookies. The type of news or information that a user wants varies from time to time. Since it is very difficult to assess the user needs, the presentation style does not impress all users. Users may not pay subscription, if there is any, since they already pay for their internet connection as well as for the telephone.

An important issue that online newspapers face is the culture of the Internet itself. Unlike the printed media, which is operated by commercial enterprises, the Internet is decentralized, open, autonomous and anarchic. This contrast and the printed medium's dislike of change have left Web critics to have the opinion that newspapers do not belong to the Internet at all. It is easier to pick up a newspaper, glance through all the headlines, pictures and articles and get a limited idea of the events and news in the world. To read an online newspaper or a magazine, a user has to turn on a computer, connect to the Internet, google for the newspaper he wants and scroll and click through all the articles and photos.

The commercialism of a newspaper and the Internet culture have together created a financial problem for online newspapers. Big advertisers are afraid to put their product online because there is more competition to deal with. Local advertisers are afraid to put their product online because they do not require their publicity through out the world as their product may be limited to the concerned area only.

When the news incorporates links paid for by corporate clients, the credibility gets jeopardized. Newspapers are also afraid that if their advertisers put their product online, then they may lose their secrecy. Moreover, various undue links to vague and vulgar web sites including porn sites are serious problem as they give wrong impression to younger generation and sometimes, even to children.

ONLINE PUBLICATIONS OF NEWSPAPERS IN INDIA

A large number of newspapers are being published online in India to facilitate the general mass and/or a special group of people who need to access the information for their specific purposes. The following is an abridged list of leading newspapers which are being published online.

LEADING ONLINE NEWSPAPERS

1. 'Amar Ujala' a Hindi daily (<http://amarujala.com/>)
2. 'Daily Hindi Milap' daily newspaper (<http://www.hindimilap.com>)
3. 'Indian Express' a English daily (<http://www.indianexpress.com/>)
4. 'Akila Daily' a Gujarati daily (<http://www.akiladaily.com/>)

5. 'Andhra Bhoomi' a Telugu daily (<http://www.deccan.com/bhoomi/>)
6. 'Andhra Jyoti' a Telugu daily (<http://www.andhrajyoti.com/>)
7. 'Andhra Prabha' a Telugu newspaper (<http://www.andhraprabha.com/>)
8. 'Business Line' a financial daily (<http://www.hindubusinessline.com/>)
9. 'Business Standard' a business related daily (<http://www.business-standard.com/>)
10. 'Daily Excelsior' (<http://www.dailyexcelsior.com/>)
11. 'Dainik Bhaskar' a Hindi daily (<http://www.bhaskar.com>)
12. 'Dainik Jagran' a Hindi daily (<http://www.jagran.com/>)
13. 'Deccan Herald' (<http://www.deccanherald.com/>)
14. 'Dinakaran' a Tamil daily (<http://www.dinakaran.com/>)
15. 'Dinamani' a Tamil daily (<http://www.xpres-net.com/dinamani/>)
16. 'Gomantak Times' an online newspaper from Goa (<http://www.gomantaktimes.com/>)
17. 'Hindimilap' a Hindi daily (<http://www.hindimilap.com/>)
18. 'Hindustan' a Hindi daily (<http://www.hindustantimes.com/>)
19. 'Kannada Prabha' a Kannada daily (<http://www.kannadaprabha.com/>)
20. 'Kerala Express' a Malayalam daily (<http://www.keralax.com/kxp/kxp.htm>)
21. 'Lashkar' an Urdu weekly newspaper (<http://www.lashkar.com/>)
22. 'Lokmat Times' a Marathi daily (<http://www.lokmat.com/>)
23. 'Malayala Manorama' a Malayalam daily (<http://www.malayalamanorama.com/>)
24. 'Mathrubumi' a Malayalm daily (<http://www.mathrubhumi.com/>)
25. 'Mid-Day' daily from Mumbai (<http://www.mid-day.com/>)
26. 'Naidunia' a Hindi daily (<http://www.naidunia.com/>)
27. 'Navbharat Times' a Hindi daily (<http://www.navbharattimes.com/>)
28. 'Panjab Kesari' a Hindi daily (<http://www.panjabkesari.com/>)
29. 'Prajavani' a Kannada daily (<http://www.prajavani.net/>)
30. 'Pratidin' a Bengali daily (<http://www.pratidin.com/>)
31. 'Rajasthan Patrika' internet edition (<http://www.rajasthanpatrika.com/>)
32. 'Rastriya Sahara' a Hindi daily (<http://www.RashtriyaSahara.com/news/>)
33. 'Sambad' a Bengali Internet newspaper (<http://www.sambad.com/>)
34. 'Sambhaav' a Gujarati Daily (<http://www.sambhaav.com/index.html>)
35. 'Sandesh' a Gujarati daily (<http://www.sandesh.com/>)
36. 'Suprova' a Bengali daily (<http://www.suprova.com/>)
37. 'The Asian Age' (<http://www.asianage.com/>)
38. 'The Assam Tribune' (<http://www.assamtribune.com/>)
39. 'The Deccan chronicle' (<http://www.deccan.com/>)
40. 'The Economic Times' a financial daily (<http://www.economicstimes.com/today/pagehome.htm>)

41. 'The Hindu' (website <http://www.the-hindu.com/>)
42. 'The Hindustan Times' a English daily (<http://www.hindustantimes.com/>)
43. 'The Kashmir Times' a Jammu and Kashmir newspaper (<http://www.kashmirtimes.com/>)
44. 'The Munsif Daily' an Urdu daily (<http://www.munsifdaily.com/>)
45. 'The New India Express' English daily (<http://expressindia.com/>)
46. 'The Poineer' an English daily (<http://www.dailypioneer.com/>)
47. 'The Siasat Daily' an Urdu daily (<http://www.siasat.com/>)
48. 'The Statesman' a English daily (<http://new.thestatesman.net/>)
49. 'The Telegraph' a leading English daily in the North Eastern part of India (<http://www.telegraphindia.com/>)
50. 'The Times of India' (<http://www.timesofindia.com/>)
51. 'The Tribune India' (<http://www.tribuneindia.com/>)

CONCLUSION

Whereas a newspaper might have previously been competing on the news stand against a handful of other newspapers, on the Internet it is competing against hundreds, perhaps thousands of other newspapers including various news magazines and millions of independently initiated documents and multi-media presentations, each of which has varying relevance to its separate readers' needs. Online newspapers and magazines have brought the whole world at reader's doorsteps. These online editions fulfill the law, 'Information for all' by their universal availability and access.

The Internet and digital technologies offer the promise of greater dissemination of information and technology globally. All human activities are dependent on information. Information has become so vital for all processes, without which any human activity can hardly take place. This technology has made a deep impact on all information-related products and services including the newspapers and other relevant publication. The increasing user population in the Net has attracted the traditional newspapers and magazines to publish their online editions. In India, this online medium has a long way to go when compared to the western countries. It should be designed in such a way that users of all age and interests get attracted by them and derive benefit..

REFERENCES

1. *News Resources on the World Wide Web (Database. Vol-19. No-1. Feb/March-1996)* by Greg R. Notess. pp.12-20.
2. *Current News Services on Compuserv- Fast and Affordable (Database. Vol-19. No-1. Feb/March-1996)* by Amelia Kassel. Pp.21-30.
3. *Where in the World is the New York Times? A Searcher watches and Hopes (Database. Vol-19. No-3. June/July-1996).* Pp.28-41.
4. <http://www.cjr.org/year/96/2/tour.asp>
5. <http://ajr.newslink.org/mmcol.html>

6. <http://www.sentex.net/~mmcadams/invent/invent1.html>
7. <http://telecomblogs.in/wp-content/uploads/2010/01/InternetUsers-in-India.jpg> (growth chart)
8. <http://www.internetworldstats.com/top20.htm>
9. <http://www.scribd.com/doc/21485962/Internet-penetration-in-India-2008-to-2013>
10. <http://trak.in/tags/business/2007/08/14/india-internet-usage-statistics-analysis-future-indian-internet/>
11. <http://timesofindia.indiatimes.com/biz/india-business/India-to-have-3rd-largest-number-of-Internet-users-by-2013/articleshow/4821630.cms>
12. <http://www.medianama.com/2008/12/223-india-has-81-m-internet-users-please-make-changes-to-your-powerpoint-slides/>
13. <https://drtc.isibang.ac.in/bitstream/handle/1849/108/Bf.pdf;jsessionid>
14. <http://jcmc.indiana.edu/vol4/issue1/neuberger.html#ADVANTAGES>

NEWSPAPER SECTION OF THE NATIONAL LIBRARY KOLKATA: PHYSICAL PRESERVATION AND ACCESS TO NEWS

Gopa Ghosh

Nivedita Bhattacharya

and

Prof. (Dr.) Biplab Chakrabarti

ABSTRACT

This paper describes the historical background of the newspaper collection of National Library, Kolkata. It depicts the current status of this collection with the present set up of its organization. The process of physical preservation of these valuable heritages has also been described here. This collection includes newspapers starting from late 18th century to the current date. The Indian Mirror, the Pioneer, and the Hindu, Patriot etc. are some of the old titles, which have already been microfilmed. In addition to this, today, there are many newspapers of different Indian languages like Bengali, Kannada, Assamese, Gujarati, Oriya, Tamil, Telegu etc. in the Library. It provides a close view on the issues regarding the use of this collection. The present challenges faced by the Library and its future plans in the present context of information society have been discussed here.

Keywords: National Library of India, Newspapers, Acquisition of Newspaper, Preservation of Newspaper, Conservation of Newspaper

INTRODUCTION

A newspaper may be a weekly or daily publication. Primarily it contains news based on the day to day happenings. Different kinds of information and advertisements are also present there. General-interest newspapers often feature articles on political events, crime, business, art/entertainment, society and sports. Most traditional papers also feature an editorial page containing columns that express the personal opinions of writers. Supplementary sections may contain advertisements, comics, and coupons (<http://en.wikipedia.org/wiki/Newspaper>). A wide variety of features like editorial columns, criticism, persuasion, obituaries, entertainments such as crosswords, Sudoku and horoscopes; weather news and weather forecasts; columns for advice, gossip, food etc. occupy too much space in any kind of newspaper. Critical reviews of movies, plays and restaurants; classified advertisements; display advertisements, editorial cartoons and comic strips also appear in regular issues of newspapers. These are the good sources to spend leisure hours comfortably for some people on one hand and to gather updated knowledge for others. Whatever be the purpose, newspapers are a part and parcel of our daily lives.

Reporters collect the news items from the reality and write the articles for the newspapers. Photographers shoot pictures to accompany the news items and graphic artists contribute charts and diagrams in relation to the eventual information. Editors assign reporters to narrate stories on the basis of their collected news items. Editors also check over those stories, write headlines for them, determine where they will be placed in the newspaper and then design the layout of the paper i.e., the arrangement of stories, photographs and art on each page. An editor-in-chief or an executive editor usually supervises the paper's news staff. The newspaper's publisher has overall control of its business and news operations.

The modern newspaper is a European invention. Missionaries and other foreigners introduced the modern newspapers in China in the nineteenth century. The oldest direct ancestors of the modern newspaper are handwritten news sheets which were circulated widely in Venice in the sixteenth century. Most of the cities like Venice played a major role in the early history of the newspaper. At that time Venice was a center for trade and therefore for information. The Venetian news sheets, known as *avisi* or *gazette*, were filled with information on wars and politics in Italy and Europe. These were distributed weekly as early as 1566 and these were seen also in London far from Venice (www.nyu.edu/classes/stephens/Collier's_page.htm). The style of journalism started at that time were also very constructive, comprising of short sets of news items, forwarded from a particular city, written under the name of that city and the date on which these were sent. This was the style that would be used in most early printed newspapers.

In India besides the several newspaper libraries, National Library is an important institution which holds copies of all the newspapers published in India and also purchases copies of selected foreign newspapers.

THE NATIONAL LIBRARY OF INDIA

The National Library of India is the apex body in the library system in India. Serving as a public library this is a permanent repository of all documents published in India in whatever language. This is the largest library in the country and an institution of national importance under the 'Ministry of Culture, Government of India'. Under the Delivery of Books and Newspapers (Public Library) Act, 1954, the National Library is entitled to receive one copy of each publication published in the country. It is also a repository of the United Nations publications as well as official documents of selected foreign countries. The Library has been receiving documents of more than 40 foreign languages through exchange and purchase. Besides these a good number of selected documents in the English language on various subjects are purchased here every year. The Library also holds a considerable number of electronic documents, some of which are received along with their hardcopies through the above mentioned act, some through purchase and some are generated within itself through scanning of old brittle documents. This Library has also a collection of rare documents. It collects, disseminates and preserves all these materials. It is situated on the scenic 30 acres Belvedere Estate, in the city of Kolkata, formerly Calcutta City.

The Calcutta Public Library was established in 1836 and the Imperial Library was formed in 1891 by combining a number of Secretariat Libraries. Lord Curzon, the then Governor General of India, decided to amalgamate the rich collection of both of these libraries to make Imperial Library as 'a library of reference, a working place for students and a repository of material for the future historians of India, in which, so far as possible, every work written about India, at any time, can be seen and read' (www.nationallibrary.gov.in). The library called Imperial Library was formally opened to the public on 30th January 1903 at Metcalfe Hall, Kolkata. John Macfarlane, the Assistant Librarian of the British Museum, London, was appointed as the first Librarian of the Imperial Library. After his death, the famous scholar and linguist Harinath De took over the charge of the library. After his death J. A. Chapman became the librarian. Mr. Chapman showed keen interest in the affairs of the library and tried hard to improve its status. After his retirement, Khan Bahadur M.A. Asadulla was appointed as the librarian and he continued as the librarian till July 1947 (Kasavan 1961).

After independence the Government of India changed the name of the Imperial Library to the National Library, with an enactment of the Imperial Library (change of name) Act 1948 and the collection was shifted from the Esplanade to the present Belvedere Estate. On 1st February 1953 the National Library was opened to the public, inaugurated by Maulana Abul Kalam Azad. Sri B. S. Kesavan was appointed as the first librarian of the National Library. In 1954 Delivery of Books Act was enacted by the Government of India in order to develop the collection of the National Library and to achieve its goal of serving the nation.

GLIMPSES OF NEWSPAPER COLLECTION OF THE NATIONAL LIBRARY OF INDIA

The Library has a rich collection of late 19th and early 20th century newspapers including "Hicky's Gazette", the first newspaper in India. There are 66 important English titles of the 18th, 19th and early 20th centuries in the Esplanade Reading Room of the Library. Some of them are Bombay Courier, Bombay (1794-1827); Sun, London (1816-1834); Englishman and Military Chronicle, Calcutta (1833-1934); Pioneer, Allahabad & Lucknow (1867-1988); The Statesman, Calcutta (1875-1988); The Times, London (1881-1971); Indian Mirror, Calcutta (1883-1923); Amrita Bazar Patrika, Calcutta (1886-1988); Hindu Patriot, Calcutta (1854-1922); New York Times, New York (1949-1967) etc. Nine important Bengali titles of the 20th century are present in the Esplanade Reading Room collection of which Vandemataram, Calcutta (1901-1908); Bangabasi, Calcutta (1929-1933); Ananda Bazar Patrika, Calcutta (1948-1988); Dainik Basumati, Calcutta (1954-1970), Jugantar; Calcutta (1954-1988) are most prominent. Nine important Hindi titles of the 20th century are present in this collection of which Lokmanya, Calcutta (1930-1965); Hindustan, Delhi (1948-1988); Aaj, Benaras (1948-1988); Dainik Viswamitra, Calcutta (1948-1988); Sanmarg, Calcutta (1957-1988) are the most prominent titles. Marathi title like Kesari (1953-1963), Malayalam title like Malayala Manorama (1959-1963), 33 Urdu titles like Rozana Hind, Calcutta (1950-1964); Azmal, Bombay (1957-1964); Quami Awaz, Lucknow (1957-1964) etc. took an important part of

this collection. All of these are bound volumes. There are an approximate total of 7812 bound volumes of newspapers of 119 titles in the Newspaper Stack and Reading Room at 5, Esplanade East, Kolkata (Esplanade Reading Room).

Many old important titles of the 19th and 20th centuries are present also in microfilm form like Amrita Bazar Patrika (10 July 1947-30 Oct 1947), The Hindu Patriot (April, 1854-August, 1923), The Indian Mirror (Jan, 1883-Dec, 1923), Bengal Harkuru and the Indian Gazette (July 1828 to May 1866 with gaps), Calcutta courier (1832-1842) etc. There are total 42 titles of microfilm copies of old newspapers in the Library. Approximately 5000 rolls are present there.

At present almost all the newspapers published in India are present in the collection. The total number of English titles of Indian newspapers received by the Library is 98, of which The Statesman, Hindustan Times, The Telegraph, Economic Times etc. are the prominent titles. Some selected foreign newspapers are also purchased every year (e.g. The Times; London, Guardian Weekly, Oleserver, Time Educational Supplement etc.)

Newspapers in regional languages like Tamil, Telegu, Punjabi, Gujarati, Oriya, Kannada, Assamese are also present in the collection of the National Library. Current regional language newspapers like Samaj, Dharitri, Anupam Bharat etc. in Oriya, Maharashtra Times, Lokmat etc. in Marathi, Gujarat Samachar, Sandesh etc. in Gujarati, Dinatandi, Dinakaran etc. in Tamil, Enadu, Vartha in Telegu, Prajavani, Udayavani etc. in Kannada, Malayalam Manorama in Malayalam language, Dainik Janmabhumi, Dainik Assam in Assamese, Ajit daily, Punjabi tribune etc. in Punjabi Ananda Bazar Patrika, Ajkal, Pratidin etc. in Bengali are present in the collection.

Besides these, some titles published in other languages like The Siasat daily, Milap etc. in Urdu, Sanmarg, Dainik Tribune, Janasatta etc. in Hindi are also present in the collection of the Library. Now there are 43 titles of newspapers of different languages (excluding English) are received by the library.

IMPORTANCE OF THE COLLECTION

The news becomes history. To trace the history, people look for the newspaper archives. The old stock in the Esplanade Reading Room helps us to know many facts about Indian culture, tradition and happenings. They throw light on the British rule in India and the period of freedom movement in India. The collection is quite important for researchers. Actually the newspaper can visualise all the past events. Many important and rare news items can be found out from these old collections. So this collection holds an important place in the society. The current English newspaper collection is also very useful tool for reference work for people from different vocations. The collection of newspapers in the regional languages of India is also attracts many researchers and scholars of varied interests. Many foreign researchers also come to the National Library to gather old news items.

ORGANISATIONAL SET UP OF THE NEWSPAPER COLLECTION

More than one division are involved in the organisation of the newspaper collection of the National Library. These are the English Serials Division, the Newspaper Section, the Esplanade Reading Room, the Rare Book Division and the language divisions.

The shifting of library from Esplanade premises to Belvedere was started at the end of 1948. The main hall of the old building was continued as an extension of reading rooms of the library for reading loose periodicals. In early 1960, the bound volumes of newspapers were shifted to the Esplanade premises and then that hall was renovated in order to become a full-fledged newspaper reading room (Kasavan 1961). The area of this reading room is 1781 sq ft and the number of reader's seat is 22 (Majumder 1987).

At present the Newspaper Section of the library only stacks current newspaper. The Newspaper Section is placed at the 6th floor of the Annexe Building. This division was started from 1972 (National Library 1972). Before that loose issues were stacked in the main stack. English newspapers are organised by the English Serials Division. This division is placed at Bhasha Bhavan. Newspapers of regional languages are organised by respective language divisions, placed at the Bhasha Bhavan also. The old bound volumes of newspapers are organised by the Esplanade Reading Room. Microfilming of old newspapers are done in the Reprography Division. Microfilm rolls are stored and organised by the Rare Book Division, placed at the ground floor of the Annexe Building.

ACQUISITION OF NEWSPAPER

According to the Delivery of Books (Public Libraries) Act, 1954, the National Library is entitled to get one copy of every book published in India free of cost. This act was amended in 29 December, 1956. According to the Delivery of Books (Public Libraries) Amended Act, 1956, National Library is entitled to get newspapers also free of cost. Before 1956 the newspaper collection was based on the collection of the Imperial library, the means of development of collection was mainly through gift

At present all the newspapers of Indian languages are received in their respective language divisions through the above mentioned act. English newspapers published in India are received by the English Serial Division also through the act. The English newspapers published in foreign countries, are purchased and acquired by the English Serial Division also.

ORGANISATION OF THE NEWSPAPER COLLECTION

Old bound volumes of newspapers up to 1988 are stacked up in the Esplanade Reading Room. All of these volumes have been catalogued. The volumes are shelved according to call number. All the current issues are catalogued by the English Serial Division according to AACR-II. Loose issues are recorded in kardex. These will be converted into MARC-21 through retro conversion project very soon. This project is continuing. The Newspaper Section at the Annexe

Building of the Library is the stack of newspapers published after 1988. All the loose issues are shelved alphabetically according to the titles.

Microfilm rolls of old important newspapers are processed in the Rare Book Division. These are stored within the cabinet in a systematic way.

CONSERVATION OF NEWSPAPER

B. S. Kesavan developed a preservation unit in 1941. Before 1941, the publications in the Imperial Library were bound at the Government of India Press. After that, private contractors were engaged for this job. In 1959, preservation unit was started. It comprised of preparatory unit, binding unit and mending unit. But it was observed that mending, treating, laminating and binding were undoubtedly laborious jobs and difficult tasks considering the size of the collection. As an alternative, in order to microfilm all the important rare and brittle publications, which are unfit for use, a reprography division came into existence in 1970 with a number of expensive and sophisticated cameras, several Xerox machines, microfilm readers, reader-printers, microfilm copier, and other auxiliary equipments.

Again lamination and binding of newspapers were also found to be problematic issues due to the quality of newsprint and the large size. Deacidification is also not the successful way of conservation of newspaper at all. Some types of acid in newsprint (wood-pulp) paper cannot be removed by any means. Acid was used in the manufacture of wood pulp paper and it does not matter how much the paper was washed after it was made, or whether the paper is a year old or a century old, some of that acid still remains breaking down the fibers of that paper. As a result browning, brittleness and chipping of newsprint paper have been seen. For long-term preservation of newspapers printed on wood-pulp paper it is necessary to deacidify them. This involves soaking, spraying or brushing with a deacidification solution. This solution will wet the paper, and wet newsprint is extremely fragile and very easily torn.

So the library started microfilming files of national dailies, old and rare newspapers with other rare books of its stock. Master negatives are preserved for archival purpose. Duplicate negatives are made from those master negatives to produce service copies for library users. Master negatives are stored separately.

Preventive Measures

Newsprint wasn't designed to last an eternity, but it can be made to last longer by keeping it away from light, heat and moisture. To achieve this goal following measures have been taken by the National Library Authority.

- 1) Dusting of newspapers through cleaning and then blowing with vacuum cleaner is done in the newspaper stacks at regular intervals.
- 2) Spraying with PIP (petroleum based chemicals like pyrethrum) is done in newspaper stack of the Annexe Building and in the stack of the Esplanade Reading Room at the regular intervals of 2 months to remove stain from the paper.

- 3) Normal room temperature causes rapid brittleness and degeneration of the newspaper. Humidity is also a great factor to affect any type of library materials in the tropical country like India. So regulated temperature and humidity are ideal for any type of library document. Ideal storage condition of 18°C to 20°C temperature and 50% to 55% relative humidity through air-conditioning is maintained here. The Annexe building, where loose issues of newspapers are stacked and where the Rare Book Division is present, is centrally air-conditioned.
- 4) The stack of old bound volumes of newspapers and the reading room at Esplanade will be shifted very soon to the new building, Bhasha Bhavan, which is also centrally air-conditioned.
- 5) Loose issues of newspapers are kept within polythene bags for protection against sunlight and fluorescent light, both of which are harmful for them.
- 6) In the Rare Book Division microfilm rolls are stored in boxes, which are kept in cabinets.
- 7) Fire prevention system is present in all the buildings.

ACCESS TO THE COLLECTION

Well maintained catalogues help the users to access the collection. In the Rare Book Division they can consult the catalogue of microfilmed newspapers. Users can read the microfilms using microfilm readers. In the Esplanade Reading Room they can get access to the catalogue of bound volumes of old newspapers. Such a separate reading room of newspaper is very much helpful for the users and it is also very convenient for the library professionals to serve the users. In the Annexe Reading Room the users can get access to the catalogue of loose issues. Users can get copies through microfilming or xeroxing. They can also get digitised copy using their own digital camera after getting necessary permission. These facilities are available in all the reading rooms.

USERS OF NEWSPAPER ARCHIVES

Researchers, politicians, students, journalists and historians mainly use the collection of old and current issues of newspapers. Average number of users at the Esplanade Reading Room is 267 per month and average number of volumes supplied to the users is 283 per month according to the in-house records maintained there. Most of them are researchers and educationists. They use mainly Amrita Bazar Patrika, The Statesman, Hindu Patriot, Ananda Bazar Patrika, The Times of India, Morning Chronicle, and Jugantar. Primarily they search for information of before 1947. Average number of users for current issues of newspapers is 174 per month and average number of loose issues supplied to the users at the Annexe Reading Room is 578 per month according to the in-house records maintained there. People of variant professions also consult the old newspapers in order to search their family history. There are also a good number of users of microfilm copies of newspapers in the Rare Book

Division. This is 105 in average per month. Many people from all over India and abroad visit the National Library to consult many old newspapers.

AVAILABILITY OF INFRASTRUCTURE

In order to organise and maintain such a big collection of Newspaper, following infrastructure is available at present in the Library.

- **Manpower:** In the English Serial Division 2 technical persons (library professionals) are engaged in procurement and processing the current titles of newspapers. In the Newspaper Section 1 technical and 2 non-technical persons are engaged in organisation of the stack of loose issues of newspapers. In the Esplanade Reading Room 1 technical and 2 non-technical persons manage the collection of old bound volumes of newspaper. Only one person is engaged in microfilming work in the Reprography Division. In the Rare Book Division 2 technical persons and one non-technical person are engaged to manage the microfilmed newspaper along with the whole collection of the Division.
- **Materials:** Newspapers are shelved at steel shelves. Kardex is available for recording loose issues of newspapers. Catalogue cards for bound volumes and the microfilm copies of newspapers are kept in catalogue cabinets. 3 Microfilm cameras (Imtec, fuji, zeiss) are present in the Reprography Division for microfilming the entire brittle documents including newspaper. 5 Microfilm readers are available in the Rare Book Division. 1 xerox machine is available in the Annexe Building Reading Room. 1 microfilm copier is present at the Reprography Division. Computer terminals are present in all the divisions which are connected through Local Area Network (LAN) within the Library premises. Upgradation of the existing set up of information communication technology is currently going on.

CHALLENGES

In the present context of development of information communication technology and its application in storage and retrieval of library materials, this library has to face some challenges.

- **Incompleteness in the collection:** In case of newspaper collection, it is very difficult to recover the gap issues. The delivery of books act has no provision for punishing the defaulters, who don't obey the act. It should include electronic documents under its purview, so that this library can get the digital copies of newspaper also. The Present act should be amended as soon as possible. To complete its collection the National Library should obtain microfilm copies from other libraries through resource sharing.
- **Scattering of collection:** All the newspapers should be available to the users at one place. It will be convenient for them.
- **The National Library should procure more foreign newspapers.**
- **The National Library should digitise all the newspapers.**

- Shortage of Manpower: There is a serious shortage of manpower in all the divisions of the Library.
- Documentation Work: Indexing should be done for important articles, so that users can get relevant information readily. But there is a serious shortage of competent professionals, who can perform the indexing work.
- Microfilming of regional language newspapers should also be done on the basis of urgent need.

FUTURE PLANS

Digitisation is an important aspect in the preservation of documents for posterity and the National Library of India is no exception in this matter. Though there are a number of challenges, still the staff members are trying hard to preserve the library materials for the future. For this the authority of the National Library of India has thought of the following future plans:

- 1) Microfilming of all the newspapers at random through private contractor.
- 2) Reproduction of digital copies from the microfilms.
- 3) Purchasing of digital copies from newspaper agencies.

CONCLUSION

Newspapers are basically self-destructing things, printed with poor quality ink and paper. But it is an indispensable source of all kinds of information on day to day life within a country as well as though out the world. Without them modern society cannot survive. The greatest part of general information is found in newspaper only. So there is a wide variety of users engaged in searching newspaper to get their relevant information in the Library. Newspapers are an important element for the society, as they keep us updated about everyday happenings. At the age of explosion of electronic content and changing information behaviour of users, the National Library should adopt the digital trend in its collection, organisation and services. To meet future demands of the readers, the Library should implement the decisions of the digital storage of this newspaper collection most urgently. Despite the above mentioned challenges, the National Library should build up a digital infrastructure to enable the whole newspaper collection digitised and to make it accessible to the users more conveniently. Many of us think that preserving newspaper is not that essential as news items are relevant only when they are current. But when we want to look back into history then these old newspapers become our utmost necessity. The National Library of India is doing this task of preserving the newspapers from the good old days. The collection is very helpful and important for the society as a whole, as people from far-off land often visit the Esplanade Reading Rooms and Annexe Reading Rooms to consult them. These collections are a treasure and let's hope that they serve the society forever.

ACKNOWLEDGEMENT

1. Mr. Aresh Kumar Ghatak, National Library, Kolkata.
2. Mr. Jagmal Singh, National Library, Kolkata.
3. Mr. C.K. Islam, National Library, Kolkata.
4. Mrs. Malabika Ghosh, National Library, Kolkata.
5. Mr. Pratap Maity, National Library, Kolkata.

REFERENCES

1. (Kesavan, B. S.). 1961. *India's National Library* Calcutta: National Library. pp.87
2. (Majumder, Uma). 1987. *India's National Library: systematisation and modernisation*. Calcutta: National Library. pp. 226. National Library. 1972. Annual report, 1971-72. Calcutta: National Library. pp.5-8
3. Newspaper-Wikipedia, the free encyclopedia
<http://en.wikipedia.org/wiki/Newspaper> (accessed on 31.1.10)
4. The National Library, Kolkata www.nationallibrary.gov.in (accessed on 31.1.10)
5. Stephens, Mitchell. History of Newspapers
[www.nyu.edu/classes/stephens/Collier's page.htm](http://www.nyu.edu/classes/stephens/Collier's_page.htm) (accessed on 31.1.10)

BUILDING A WEB BASED DIGITAL REPOSITORY OF IMAGES: CASE STUDY OF THE HINDUSTAN TIMES

Pratibha Kaushik

ABSTRACT

Print and Electronic media is a massive industry. Rapid advancements in the information and communication technology affect this sector more than others. Digital repository of images requires a strategy for the storage of large quantities of data, which increases dramatically when dealing with high resolution images. The Hindustan Times (English daily) is one of India's largest circulated dailies which offers holistic coverage of India and the world. Head office of the newspaper is situated in Delhi and editions are published from Delhi, Mumbai, Mohali, Patna, Lucknow, Kolkata Jaipur and Bhopal etc. We have countrywide network of journalists and different supplement editions with a daily circulation like HT City, Real Estate, Live, Horizons, and HT Business etc. To fulfill the needs and requirement of Images of our users and different centers, we have established a web based Digital Repository. This paper describes the practical experiences gained during the establishment of the repository. It also discusses about software installation, hardware, architectural overview, processing of images, adding metadata, indexing and standard etc.

Keywords: Digital repository, Archiving of Images, Indexing, Metadata, Newspaper library.

INTRODUCTION

The Information and Communication Technologies have transformed the way information is generated, preserved, archived and made accessible. With the technologies providing such unprecedented access to digital resources and information services, Digital Libraries and Institutional Repositories are now worldwide success. They have opened up enormous opportunities to revolutionize access to scientific and technological data and knowledge generated by academia, research institutes, government/non government organizations and the corporate sector. Libraries, the world over, are now looking forward to setting up Digital Libraries using commercial and open source software and provide open access for resource sharing. Developing such Digital Repositories is a tremendous challenge, both for government institutions and the corporate sector.

INTRODUCTION TO THE HINDUSTAN TIMES

The Hindustan Times is commonly called as 'HT'. Time was when the reader of the Hindustan Times was assumed to be a resident of Delhi. Today the paper has eight editions across the country. With advances in communication, every region finds space in every edition, making it a truly national newspaper. But the story of the Hindustan Times began in 1924. The paper, which is among the

oldest of all the existing English dailies of Delhi, was born in the Old Delhi railway station area and it was founded by Master Sunder Singh Layallpuri, the founder father of the Akali Movement and the Shiromani Akali Dal in Punjab. Mahatma Gandhi inaugurated the newspaper. The Hindustan Times had one edition for the first 60 years of its existence.

From hand-made typeset lino-printed pages to 'read anywhere' electronically made anytime accessible newspaper, the Hindustan Times has come a long way. The paper is now available to readers across the globe on e-readers like Kindle, uploaded every at 6 a.m.

The Hindustan Times (HT) today has become a leading newspaper in India because of the authenticity of its news - the newspaper has become a market leader for English papers in north India. It has proved its nation wide reach in India. Leadership through quality and innovation is the hallmark of the Hindustan Times Media Limited. Head office of the Newspaper is situated in New Delhi and paper editions published from New Delhi, Mumbai, Lucknow, Patna, Bhopal, Mohali, Jaipur and Kolkata. HT Delhi edition usually contains 22-24 pages of main newspaper excluding a daily lifestyle supplement called HT City and other supplements on weekdays. On Tuesday HT Delhi carries a Job supplement called 'Power Jobs' and an education supplement HT Horizon on Wednesday. Real Estates section called HT Estates and Luxury supplement called HT Splurge are circulated on Saturday. Every Sunday HT carries a special magazine called HT Brunch, which carries articles on lifestyle, entertainment, food, travel, fashion, health etc.

HT Next is a complete newspaper for the generation next, pursuing their education and includes everything useful that the youth want to read. It provides space for budding writers and strongly encourages participation in debates 'HT PACE' endeavours to add value to education and a new dimension to classroom teaching. PACE has made the newspaper very useful for school going children.

Hindustan Times Media Ltd. Sister publications are:-

Other sister publications of the Hindustan Times are Mint (English business daily) Hindustan (Hindi Daily), Nandan (monthly Hindi magazine for children) and Kadambani (monthly Hindi literary magazine).

The HT media group also owns a radio channel Fever 104 and organizes an annual conference which invites speakers from Indian and foreign politics, designers, businessmen, sports personality and also from the entertainment industries.

In March 2009, the Indian Readership Survey (IRS) Round I determined Hindustan Times readership to be 63.4 lakhs. HT continues to hold the ranking 2 among the Top Ten English Dailies in India.

INTRODUCTION TO DIGITAL REPOSITORY

Digital Repositories or Libraries are digitally accessible organized collection of knowledge. Such repositories are a set of electronic resources and associated technical capabilities for creating, searching and using information. In this

sense, there are an extension and enhancement of information storage and retrieval systems that takes care of the digital data in any form (text, images, sound, static or dynamic images) and exist in distributed networks. The contents of digital repositories include data, metadata, that describe representation, creator, owner, reproduction rights and metadata that consist of links or relationships to other data or metadata, whether internal or external, to the Digital Repository.

In other words Digital Repositories are electronic libraries where the information can be acquired, stored and retrieved in digital form. Digital Repositories are the group-interlinked workstations connected to the high-speed networks. A Digital Repository or Library includes a large collection of objects, stored and maintained by multiple resources such as text, documents, video, audio, images, graphics, articles, reports and books, speech, music, online tutorials, web based repositories etc.

MERITS OF DIGITAL REPOSITORY

- Provides information for 24X7.
- Promotes universal accessibility.
- Provides access to more information than physically acquired information.
- Protects rare collection/information that is rapidly deteriorating due to over use and poor storage condition.
- Provides multiple search and access.
- Provides facility for the downloading and printing.
- Save the cost and manpower required for publishing and bringing out new editions.
- Save the space, which is required for physical documents and images prints.
- Supports multimedia contents.
- Makes sharing of information resources possible.
- Helps to reach information to its users at faster rate through online communication.

BUILDING WEB BASED DIGITAL REPOSITORY OF IMAGES: CASE STUDY

The Hindustan Times, which is among India's largest circulated English dailies, offers holistic coverage of India and the world. To fulfill the needs and requirement of images to our users and different centers we have established a web based Digital Repository. HT digital repository has visual contents mainly news and feature images taken by HT staff photographers over the past many years. Since the digital operations began in India and the photographers got the digital cameras in their hands, the influx of digital images grew at a very fast pace so much so that it became very difficult for the Media Houses to handle the entire operation. With the start of digital era the helping software also came into

use such as Cumulus Software which was acquired by HT newspaper for uploading the digital images for easy retrieval and handling for editorial purpose.

Latest to the software arsenal HT purchased license to use Fotoware Suit software for managing the entire digital images of the HT Media House which also had the challenge of centralizing the entire digital repository spread over different centers of the Group. The Image Archive Section staff stationed in mainly two centers of the media house in Delhi and Mumbai started this activity of uploading the digital data after checking and completing the images with complete caption and meta data.

The digital images that were stored in the servers were numbered in hundreds of thousands. This data also included the images produced by editing and scanning the photographs over the past few decades as a result of HT copyright. The availability of information for the archived photographs was a big challenge as the necessary information was either not mentioned on the back of photographs or was not mentioned in full. This leaves the person handling such archival heritage with little option of saving the available information in the metadata. The Image Archive Section is working on the Fotoware Software and has uploaded around 4,00,000 HT copyright digital images

INTRODUCTION TO FOTOWEB

Fotoweb is a server application for publishing archives of digital assets so that users can access the archives through a web browser. It is useful for both Intranet and Internet publishing, and a single server can be used to serve both purposes. It uses the services of Microsoft Internet Information Services (IIS) to allow clients to access the system via the HTTP Protocol using Web Browser. Fotoweb consists of several parts, some of them running as 'Plug-ins' in internet information services (IIS) and others as background services that can be controlled through the Service Control Manager (SCM).

SYSTEM REQUIREMENTS FOR THE SERVER

Software

- Windows Server 2003 with Service Pack 2 (32-bit Edition)
- Internet Information Services 6.0
- Apple QuickTime 7.1.6 (full installation)
- Microsoft SQL Server (Optional) or SQL Express 2005
- SMTP Server
- Microsoft .Net Framework 2.0
- Internet connection for web installer and product activation

Hardware

- Minimum 1 2,5 GHz processor (dual-processor system recommended)
- Minimum 2 GB free RAM
- Minimum 1 GB free disk space for program installation, additional space required for data storage and web cache.

FOTOWARE INDEX MANAGER

FotoWeb is dependent upon the FotoWare Index Manager to maintain full text indexes of the images and document archives. Even though FotoWeb can publish files in folders that are not indexed by Index Manager, it is not able to search in these folders. To allow our users to have maximum benefit of Fotoweb, it is required to first install Index Manager and create an index of all your document or Images.

CLIENT REQUIREMENTS

Fotoweb 6.0 has been tested and proved compatible with the following browsers:

- Internet Explorer 7.0 (Windows)
- Safari 2.0 (Macintosh)
- Mozilla Firefox 2.0
- Opera 8.5

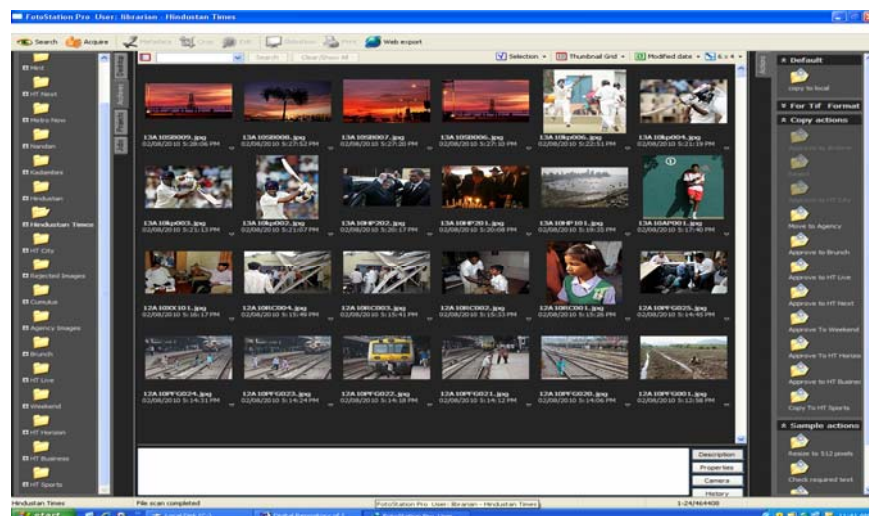
FOTOWEB DESKTOP REQUIREMENTS

- Windows XP/Vista/Mac OS 10.4(Tiger) or 10.5 (Leopard)
- 512 MB RAM
- Screen resolution of 1024x768 or higher
- QuickTime 7.1.6.
- Administrative rights for automatic installation in web browser (can also be installed through Active Directory deployment)

ARCHIVING OF DIGITAL IMAGES

HT Images Archive Section archives all the Digital Images, under HT copyright, taken by staff photographers received from all the centers through FTP server. All the digital images considered for editorial use are archived with required Metadata and checked for best quality before uploading. We are using Fotoware Fotostation for captioning and metadata.

Fotostation Overview



The FotoStation program window consists of the following parts:

Main Menu

The Main Menu together with submenus, contains all functions used in FotoStation.

Main Toolbar

In the 'Main Toolbar' you find buttons for quick access to some of the most commonly used functions. In Fotostation, hold the mouse pointer over one of the buttons in the 'Main toolbar' to view that button's tool tip. The tool tip provides information about the current button's functionality.

Image Window Toolbar

The 'Image window toolbar' contains different ways to view and sort the content of the 'Image window'. Using this toolbar you can view images as thumbnails or as detailed lists, change the sort order, change the number of thumbnails, together with many other options.

Image Window

The 'Image Window' is where the content of the current source is displayed. When the images are shown as thumbnails, this part of Fotostation looks very much like a light table.

Text Panel

Below the 'Image Window' you find the 'Text Panel'. In this part of Fotostation you find information that is saved together with the file (often called a file's metadata) together with other types of information such as file's name and size. The information is divided into four part that is (Description, Properties, Camera and History) can be viewed by clicking on the four buttons on the right hand side of the panel.

The Text Info button displays information such as the file's caption and keywords.

METADATA AND CAPTIONING OF THE IMAGES

Special emphasis is being given on providing Metadata and key words. Normally key words are such catchy strings which are mentioned in the Five Ws (who, what, when, where, why) and one H (how) of Journalism but an Image shows more than that. It is quite often mentioned that an image replaces hundreds of words. Now as a layman looks at the image he notices number of elements visible in the image. For example if the caption of the images says:

New Delhi

Feb 10, 2010

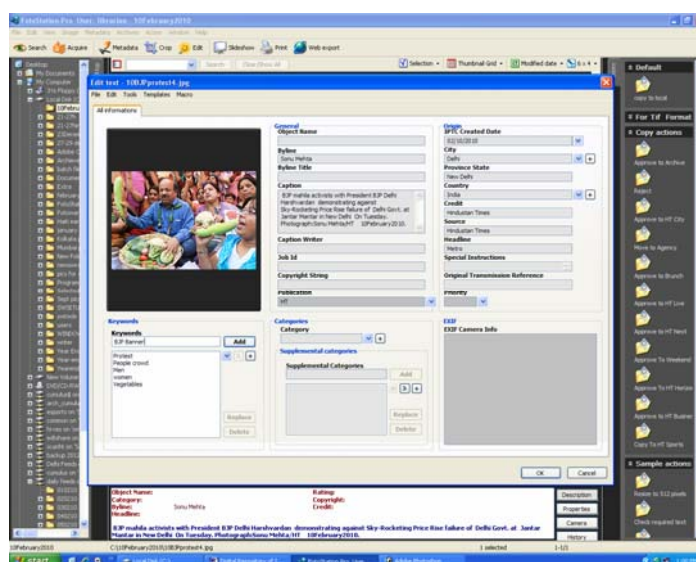
Residents of Delhi shouting slogans against the Delhi Government for its failure in controlling the price rise of food products near Jantar Mantar.

HT Photo By Sonu Mehta

Now this image shows the people shouting slogans holding banners written with slogans, showing caricatures of the leaders, women, men and even children holding vegetables, milk and other basic items.

Now the words like “children, milk, vegetables, women, men and banners, which are not mentioned in the above caption can be applied as key words strings for any possible search in future which would demand an image where the children are holding milk and women holding vegetables and men holding banners.

Example of Metadata, Caption and Keywords



Snapshot of FotoStation Metadata and Keywords

Meta Tags for Images

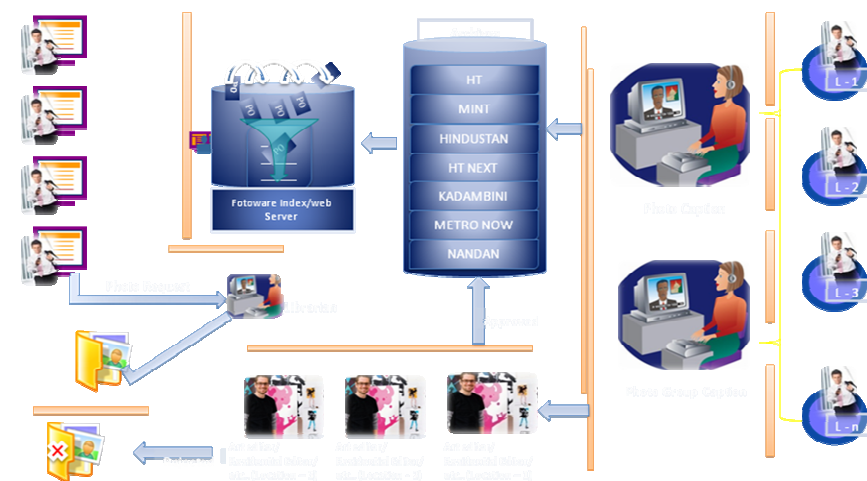
Meta Tag	Description
Image Date	Date of Image
Published date	Date on which image was published. For unpublished Images the image date is taken
Section and Suppliment	Newspaper Section like main paper and supliment i.e city, brunch, estate etc. in which image was published
Caption	Caption carried in the newspaper with the image
Description	Description of the image with all details
Keywords	Keywords follow
Category	Subject categories as per IPTC.
Source of Image	Publication Name – Hindustan Times/Hindustan/HT Next
Status	Published or Unpublished
Personality Name	Name of personality in photograph
Mood	Describes mood of the personality
Event	Describes the event of the photographs
Image Type	General news/special feature/series etc.
Colour	B&W/Colour image
Image Quality	Image size, dpi etc.
Image Place	Place where image was taken i.e New delhi, Mumbai, Patna, Kolkata, and Chandigarh etc.
Created By	Name of the Photographer
Cpoyright	Copyright Statement/ HT watermark etc.

After completing the Metadata we upload images to particular folder like HT News, City, Brunch, Business, Sports, Horizon and Estate etc. FotoStation is connected with the Fotoware Index Manager. We can understand with the help of Content and Data flow diagram given below.

Data Flow

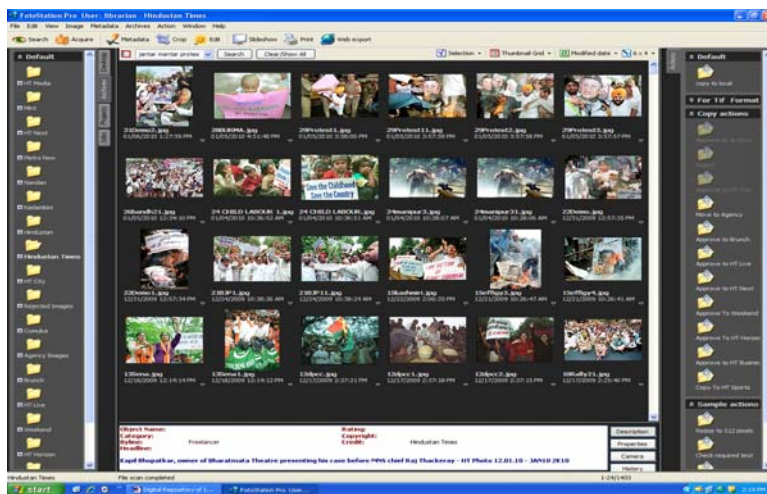


Work Flow



Retrieval and Searching of Digital Images

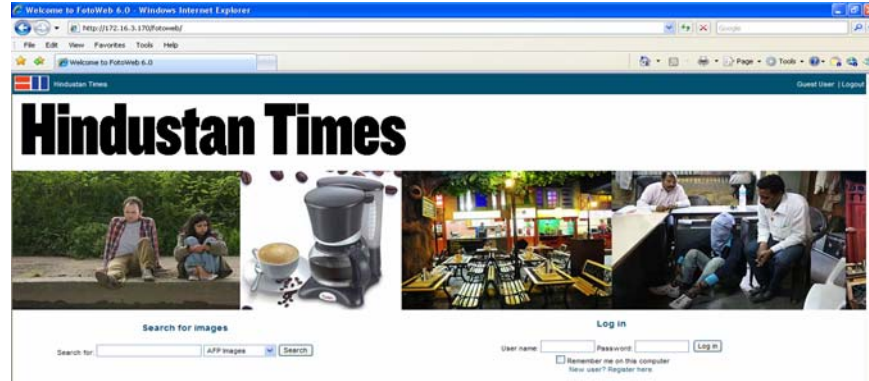
To be able to search for files, we need to have a method for adding text, so-called metadata, to each file. Without metadata, you are only able to search for file properties such as file name and file date. By adding metadata to fields, it makes easier to find images because you can search for information included in the file. With the Quick Search field you can enter search strings in a quick and straightforward manner. It provides logical operators such as AND, OR and NOT. For example if you want to search images related to Jantar Mantar protest in delhi you can type jantar mantar protest in delhi in the Quick Search



FotoWeb for User

FotoWeb is a server application for publishing archives of Digital Images so user can access the archives through a web browser. It is useful for both the

users who are using Intranet and Internet. Website is ready to use certain changes made by IT department of HT



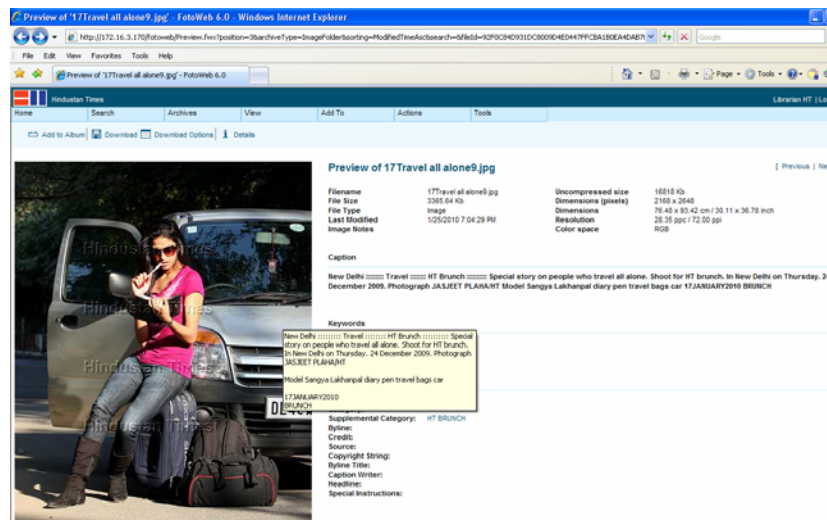
Searching the Images for News articles

Search - User can search by simple search or Advance Search.

Preview - User can preview the image and read caption related to image.

For example the story on people who travel all alone, a user may want an image of a single man or woman who has a travel bag in hand standing near a car or bus. He or she can go by simple search, or advance search by type of travel bag, car, or bus. User can see the preview of the image and if it suits to his or her story he can send the request by mail to the Image Archive Section for sending that particular image with file name. He can make Album for his requests. After that Archive Section downloads the image and sends it to the user's folder through network mapping or by FTP Server to our different centers.

All the Images are shown with HT Watermark.



Snapshot of Fotoweb 6.0 preview Windows Internet Explorer

Monitoring and Evaluation

HT IT team is checking the Fotoware server System Log and tally it on an every day basis. They analyse user reports and files. We are taking feedback of our users experience. It gives information about the user's specific requirements and how they search the images by using different key words.

CONCLUSION

Media and Entertainment is a massive industry. As per Price Waterhouse Coopers Global Entertainment and Media Outlook report the global industry is projected to grow to \$2 trillion in 2011, growing at a compound annual rate of 6.4 percent.

The Digital Repositories/Libraries are the ultimate solution for the information and data flow. That is the dream of information providers and seekers in the knowledge center system in a distributed network environment. It is also clear that Image Archive Section or Photo Library in Media centers are required to perform all the Image Management tasks in the Digital Era. Librarian and person dealing with images will have to equip themselves with capabilities to link with global trends for the ultimate benefit of information seekers. Librarian recognizes their roles in much more than custodians of the Images or Knowledge.

REFERENCES

1. Arm, William Y. Digital Libraries, Ane Books, 2005.
2. FICCI & Price Water House Coopers. The Indian Entertainment & Media Industry: Sustaining Growth, Report 2008.
3. Kumar, P.V. Future Technology : Boundary for the boundless Anna University Digital Initiation; Proceeding of MANLIBNET Conference on Business and Management Librarianship:The Decade Ahead, 2007, New Delhi.
4. Pujari, Anita and Shaikh, Fatimabee. Content Syndication business initiative by a Newspaper Library : A case Study; Proceeding of SLA-Asian Chapter(ICoASL 2008) conference on Shaping the Future of Special Libraries held at New Delhi 2008, P-69-78.
5. Rashid, Abdul. Building Digital Library : A Systematic Approach; Proceeding of SLA-Asian Chapter(ICoASL 2008) conference on Shaping the Future of Special Libraries held at New Delhi 2008, P-221-227.
6. <http://www.dlib.org/dlib/movember08/caplan/11caplan.html>
7. <http://www.dlib.org/dlib/july08/buonora/07buonora.html>
8. Fotoware Fotostation user manual.

PRESERVATION OF NEWSPAPERS IN JNU LIBRARY A CASE STUDY

Dr. M. Natarajan

and

Payel Biswas

Purpose: This paper reviews and explores the methods and techniques used for newspaper preservation in physical form and also in digital formats in Jawaharlal Nehru University (JNU) library, New Delhi. Newspapers are an important class of historical records as they give a clear view of contemporary life and events. The newspaper files preserved in libraries give valuable reference records for historical purposes. They are still the most iconic outlet for news and other types of written communication.

The study will find out the physical preservation of newspapers and newspaper clippings. It will suggest some methods and techniques, which can be used in electronic environment.

Design / Methodology: JNU library acts as a knowledge centre by acquiring different types of document collections from books, journals, periodicals, thesis, patents, standards, newspapers, CD-ROM databases, e-collections like ebooks, e-journals and e-databases. In early 1970's JNU opened its doors to teachers and students in the frontier disciplines and covering the resources for Social sciences, Humanities and Sciences. This university plays an important role in the field of socialism, mass communication and international studies. It has different categories of students from graduates, post graduates and research scholars in almost all the disciplines. Library acts also as a storehouse of documents wherein all the print collections are stored in multiple floors of the library. The JNU library procures 25 different newspapers from local, national and international countries in different languages and it acts as an important source of information for the research scholars for international studies, politics, etc. The method used for physical preservation of the newspapers is discussed in this paper. Also they archive the newspaper clippings, which are selected by special library professionals and maintained as per the subject wise arrangement. Now they have started digitizing the newspaper clippings, wherein they add the metadata for the clippings as per the subject category, authors, publishers, date of publications, year, etc. The retrieval becomes easy, as it is available in digitized environment.

Originality / Value: Preservation is the totality of the steps necessary to ensure the permanent accessibility forever. It is well known that a great deal of cultural knowledge is stored in the newspapers of yesterday. They are in popular demand and are used in university library as a tertiary source of information by the students and faculties. People interested in newspaper information benefit from these archives because the work of selecting, cutting and indexing articles is done by specialists. In order to maintain their important position in the

information market, clippings archives should be able to integrate their special skills into the technologies of the information society.

Keywords: *Newspaper, preservation, newspaper clippings, JNU, archives.*

INTRODUCTION

Local newspapers are a key resource in local studies libraries. They are used for genealogical and biographical research, and for community history (Dixon, 2001). They are accessed for school and wider research projects and also for commercial, legal and personal results. They are visual archives of the communities and people they serve. Howell (1996) described them as “Irreplaceable documents which provide vivid accounts of local places, events and people”. West (1983) maintains that “there is no source of local history as evocative of the atmosphere of any 19th century town as its local newspaper”. But Stoker (1999) pointed out that “newspapers were never intended by their producers to be a permanent means of storing textual information, and the recognition that they contain a mass of valuable information not available elsewhere, is only a comparatively recent phenomenon”. It is well known that a great deal of cultural knowledge is stored in the newspaper. Since newspaper is not always easily accessible, special clippings archives were created in the 20th century. People interested in newspaper information benefit from these archives because the work of selecting, cutting and indexing articles is done by specialists. In order to maintain their important position in the information market, clippings archives should be able to integrate their special skills into the technologies of the information society. Libraries have been undertaking digitization projects in order to provide wide access to materials in their collections This article is a case study on newspapers procured in JNU library.

DEFINITIONS

- Preservation includes managerial and financial considerations including storage and accommodation provision, staffing and policy decisions as well as the techniques and methods of maintaining materials in optimal conditions.
- Prolonging the existence of library and archival materials by maintaining them in conditions suitable for use, either in their original format or in a form more durable, through retention under proper environmental conditions or actions taken after a book or a collection has been damaged to prevent further deterioration. (The care of Fine Books, Mick Lyon Books, 1988).

JAWAHARLAL NEHRU UNIVERSITY (JNU) AND ITS LIBRARY

The JNU’s Central Library is a Knowledge Centre which has rich resources mainly in Social Sciences, Humanities and Sciences. It is a nine-storey tower building and has a carpet area of about one lakh sq. ft. It is in the hub of all the academic activities of the University. Established in 1969, it incorporates the

library of the prestigious Indian School of International Studies which was later merged with Jawaharlal Nehru University. The JNU Library is a depository of all government publications and publications of some important international organisations such as the WHO, the European Union, United Nations and its allied agencies etc. The Library has recently established a Cyber Library at the Ground Floor with 200 Computers for the students and research scholars to access the available online resources.

To meet the special needs of the visually challenged students of the University, a separate Unit named Helen Keller Unit has been established in the newly renovated Reading Hall at the Ground Floor. The Library has subscribed twenty two international online databases covering about 10,000 full text journals. Besides that, access to 4,500 full text scholarly electronic journals from 25 publishers across the globe is available, under the UGC-INFONET E-journals Consortium. Online Public Access Catalogue (OPAC) can be accessed from all the Schools / Centres under University Wide Area Network.

COLLECTION DETAILS

The Library has a total collection of 5.45 lakhs, which includes books, serials, non-book materials etc. The faculty publications have been placed at the Ground Floor. The Library subscribes to 965 journals (in print) and also receives another 148 journals (in print) by way of gift and exchange. The collection is housed subject-wise on different floors under three major streams i.e. Social Sciences, Humanities and Sciences.

COLLECTION OF NEWSPAPERS

The Central library of J.N.U. is responsible for deposit and preservation of national knowledge and information resources, collects and preserves comprehensive newspaper materials and provides library services to the researchers and scholars. It acquires many newspapers in many foreign languages (more than 25 newspapers). They mainly focus on representative daily papers of various countries including Australia, China, France, German, Japan, Kazakhstan, Russia, UK, and the USA etc. Among the list of subscription, six titles are in Japanese which is the largest in number and three titles of the USA and France respectively. They are mainly for preservation and for the use of scholars from School of Languages. It also subscribes to two titles in Urdu and Greece newspapers being a national depository. The details are as follows:

Collection of Newspapers

English	Other Languages
Asian Age, Business Standard, Dailt Assam, Decan Chronicle, Decan Herald, Economic Times,	Amar Ujala Ananda, Asomiya Pratidin, Assam, Dainik Assam, Dainik Bhaskar,

Financial Expenditure, Financial Times, The Hindu, Observer Weekly, Times Economics The Tribune.	Dainik Jagran, Desh Bandhu. Dinamoni, Enadu, Hindustan (Hindi), Inqualable Daily, Munsif Daily Roznama Rashtriya, Samaja, Sandesh, Shar Shara, Swatantra Bharat Tribune Eng and Punjabi.
---	---

USAGE STATISTICS

Since newspapers are not always easily accessible, special clippings archives were created in the 20th century. In Jawaharlal Nehru University, newspapers are at the heart of School of Information Studies, a large number of M.Phil/Ph.D students are studying in this university. They need to refer to newspapers for their research work. Media are of paramount importance for the Indian democracy. Like most Americans learn about public affairs and politics through the mass media especially newspapers and television, similar conditions apply to Indian researchers and media persons. The usage statistics of JNU newspapers utilization is given below for two years:

Table 1: Statistical Scenario Usage of newspapers

Users	2007-2008	2008-2009
Students	14.60%	17.50%
Researchers/Scholars	64.60%	70.80%
Media persons	21%	11.7%

PRESERVATION AND ACCESS TO NEWSPAPERS

Conservation Treatment of Newspapers

Newspaper issues or pages may require conservation treatment in order to preserve them as intrinsic artifacts, or for research or exhibit purposes. Conservation treatment should be referred to a professional paper conservator, since any treatment process can entail risk to both the material and the personnel involved. The conservation treatment selected will depend upon the characteristics of the individual item and its condition; testing will be done before beginning the treatment process. Acidic newsprint often requires deacidification and the deposit of an alkaline buffer (a.k.a. alkalization) to stabilize the paper. Repairs to the paper may be done using Japanese paper and wheat paste or heat-set tissue; pressure sensitive adhesive tapes are not recommended. It is important to use good quality materials that will hold up over time. In order to provide support for a fragile sheet and permit safer handling, a deacidified newspaper page can be stored or encapsulated in a polyester film sleeve or folder. It is preferable to avoid encapsulating newspapers which have not been deacidified. However, if newsprint is acidic, its

extreme fragility indicates the need for encapsulation; an alkaline buffered sheet should be placed behind the newspaper whenever possible.

Cellulose acetate lamination is not recommended for newsprint, especially for those newspapers which have intrinsic value. Lamination can extend the time over which newsprint can be actively handled, but it will also damage the paper, and is not fully reversible.

Preserving Newspaper Clippings

For decades physical form like bound form has been the only way to store and preserve newspaper collections around the world. But in case of microfilming process for newspapers the three disadvantages of non durability factor, large unwieldy size and huge storage space could all be overcome with microfilming. The first newspaper to be microfilmed was the Evening News (London), filmed in 1953 to demonstrate the viability of microfilming techniques. As early as the 1930's microphotography was recommended as a means of preserving the information available in newspapers, however, the life expectancy of film at the time was less than a generation. Today it is claimed that microfilms can last for 500 years. But in case of J.N.U. library for the huge collection of newspapers, it's not possible for a large university to maintain all newspaper collection in microfilm methods.

BINDING

One of the three copies of the newspapers collected in the JNU is provided for users, while the other is bound for preservation and other one for newspaper clippings service. All newspapers in Library are bound and especially preservation copies are bound with book binding mechanism and stored in the Deposit Building for permanent preservation. In the case of more than 10 major daily newspapers in demand, one copy is given for long term public service use and another one for permanent preservation. Some of the bound newspapers in the Deposit Building are housed in acid-free boxes to prevent physical damages. And the newspapers published before 1950, the highly valuable items are rare collections, and are specially preserved and maintained in the Deposit Building. The bound newspapers are subjected to deacidification process in the preservation office to prevent possible damage.

Many institutions will need to service and store newspapers for long periods of time prior to filming and, in some cases, may wish to store original copies on a permanent basis. Binding has been a frequently used method for organizing and storing newspaper files for many decades. Apart from the expense, binding of newspapers is often damaging to the text; creates unwieldy volumes that are difficult to handle properly; and even encourages some institutions to shelve volumes vertically, which can cause the text block to pull away from the binding. Oversize folio volumes should be stored flat. (www.loc.gov)

HOUSING

Newspapers are arranged in a chronological manner and it is retrieved by

researchers and scholars through same way. Now-a-days JNU library collects newspaper clippings from newspapers by selecting the relevant / required content for the scholars and media persons and they keep them in a chronological order.

DIGITIZATION

In Jawaharlal Nehru University, approximately 12.45 lacs newspaper clippings on area studies and basic disciplines are available for students and research scholars of the University. All clippings are arranged according to subject wise. Researchers and scholars sometime are confused to search their pin-pointed information. Digitization of newspaper clippings provides the provision for researcher to browse through Subject, Title, Author, and Name of newspaper and Date for retrieving their pin pointed information. It's bringing the each and every matter of newspaper nearer to the user, in a user friendly manner. So, on one hand our target to "save the time of the reader" is one of the important laws of library science advocated by Dr. S.R. Ranganathan. The Central library of J.N.U. started digitization process in newspapers clippings. They use the Virtua software for preparing digitized version of paper clippings. Through this proprietary software the metadata addition is possible and the process makes use of scanning methods.

BIBLIOGRAPHIC CONTROL

While often overlooked as a preservation concern, appropriate bibliographic control is an essential component to the success of any newspaper preservation programme. Complete citations assure that preserved material can be accessed by users, and that the costly duplication of preservation efforts can be avoided by others attempting to do the same work. Comprehensive bibliographic information enables one to determine who holds a given title and what issues are available or missing, as well as any comments concerning supplements, editions, or title changes. Unfortunately, few libraries, and fewer local historical societies and archives, have been able to maintain consistent bibliographic control over their newspaper collections. Any attempt at preserving newspapers without also providing bibliographic control will only exacerbate the problem, as users of poorly prepared newspaper microfilms will attest.

CONCLUSION

It should be noted here that while many current newspapers are widely available in electronic form, either via the World Wide Web or in CD-ROM format, the issue of completeness should concern anyone who would use these as a surrogate for the newsprint edition. With only a few exceptions, the newspapers available in electronic form do not always include the classified ads, legal and death notices, and other local features historians and researchers find so important. Many do not include photographs and advertising sections. If these products are allowed to serve in libraries and archives as a substitute for the

newspaper itself, then much of what is characteristic of newspapers as a tool for research is lost.

This study shows that newspapers are used as an important source of information by the research scholars and media personnel for their use. JNU library is procuring a good number of newspapers in different languages also. Even though e-newspapers are available now-a-days, full text articles may not be available for all of them. Hence JNU preserves the newspapers for better and future use by the scholars. JNU has School of Languages division, and hence it is a must for them to store and preserve the newspapers and serve the user community as per their demand. The library and information professionals are taking much care for this and also providing the e-newspapers, wherever possible.

REFERENCES

1. Dixon, D (2001). Recycling our newspapers. *Library Association Record*, 103 (5), p. 298-99.
2. Howell, A (1996). Preservation digitizing of newspapers. Conference paper, IFLA (62nd) Annual Conference, 29th August Beijing.
3. Stoker, D (1999). Should newspaper preservation be a lottery?. *Journal of Librarianship and Information Science*, 31 (3). P. 131-4.
4. West, J (1983). *Town Records*. Phillimore, Chichester.
5. <http://www.loc.gov/preserv/care/newspap.html>
provide guidance that library ease use in navigating through the various aspects of a digitization project.

BENGALI NEWSPAPERS AND THEIR NEWSPAPER PRESERVATION TECHNIQUES: AN OVERVIEW

Pritam Gurey

and

Avijit Chakrabarti

ABSTRACT

Bengal has been the proud birthplace of the Press in India. Historians record “Hicky’s Gazette” as the first newspaper to be published in the country from Calcutta in 1780. The first newspaper in an Indian language was the Samachar Darpan in Bengali. At present there are 12 Bengali newspapers available of which six are leading newspapers – Anandabazar Patrika, Aajkal, Bartaman Patrika, Sambad Pratidin, Ganashakti Patrika, and Uttar Banga Sambad – which are published both in print and electronic versions (online) from West Bengal. All the newspaper organisations have their own libraries and each of them follows their own method of preservation. These newspapers have their online version with online archives whose holding varies from nine years to one week. Online access to these newspapers poses some problems like “font”, since proper designing of ISCI compliant Bengali True Type Font (TTF) and designing UNICODE compliant Bengali Open Type Font (OTF) are challenging issues. These newspapers are loaded with local, national and international information, but finding the old information is a challenge. Unless the newspaper has been microfilmed (and often even if it has been), libraries hold very few copies and these are often not available through interlibrary loan. Even with access to a newspaper, information retrieval based on ‘keyword query’ and ‘searching’ is not possible and searching for the desired information can be a tiresome quest.

The aim of this paper is to highlight the preservation status of these newspapers and to discuss the problems of access to information both online and offline.

Keywords: Preservation, Newspaper Preservation, Bengali Newspapers, Microfilming, Digitizing.

‘A good newspaper is a nation talking to itself’ – Arthur Miller (1915-2005)

INTRODUCTION

Preservation is the oldest and most fundamental function of libraries and archives. From the ancient times the protection of cultural artifacts, including books, primary source documents and museum objects, has been described by the terms "conservation" and "preservation". Preservation is concerned with evidence embedded in the intellectual content of objects and in the objects

themselves (Wikipedia). As knowledge keepers and providers of information the social responsibility of libraries is re-emerging in the information society. By the late twentieth century, there was a trend of rapid growth of scholarly research, and the preservation of items printed on acidic paper from the mid-1800's are important recourses of their users. Newspapers provide information to readers on current events. People who read the newspaper are well aware of global and local issues as it contains news on events in distant corners of the earth. As a source of news, the newspaper acts as the most indispensable tool to those with thirst for knowledge. Chronicles record that the *Bengal Gazette* was the first newspaper to be published in the country from Calcutta in 1780. James Augustus Hicky is considered as the "father of Indian press". Bengal has been the proud birthplace of the press in India. The first newspaper in an Indian language was the *Samachar Darpan* in Bengali. The first issue of this daily was published from the Serampore Mission Press on May 23, 1818. In the same year, Ganga Kishore Bhattacharya started publishing another newspaper in Bengali, the *Bengal Gazette*. Since then, the prominent Indian languages like Hindi, Marathi, Malayalam, Kannada, Tamil, Telugu, Urdu and Bengali are published and have grown over the years. The Indian language newspapers have taken over the English press due to the marketing strategy followed by the regional papers and the growing literacy rate. Increase in the literacy rate has affected directly the rise of circulation of these regional papers. The regional papers aim to provide localised news more for their readers. Old news is often required as news becomes history, so preservation of newspapers is necessary in order to retrieve the correct information. In present times with the advent of technology, electronic form of newspapers are also available, but there are certain problems in reading the online Indic language papers because of the fonts. In this paper the preservation status of the Bengali newspapers of West Bengal is highlighted along with the problems of access to information of these newspapers both online and offline. The paper covers six leading Bengali newspaper – Anandabazar Patrika, Aajkal, Bartaman Patrika, Sambad Pratidin, Ganasshakti Patrika, and Uttar Banga Sambad which are published both in print and electronic versions (online) from West Bengal.

IMPORTANCE OF NEWSPAPER PRESERVATION

Preservation is concerned with evidence embedded in the intellectual content of objects and in the objects themselves. So the newspaper preservation is important:

- as news becomes history, and we should know our history well,
- to improve access,
- as daily important facts are found,
- Newspaper may perish but not the news they contain.

OVERVIEW OF BENGALI NEWSPAPERS IN WEST BENGAL

<u>Name</u>	<u>Print</u>	<u>On-Line</u>	<u>Website</u>
Aajkaal	✓	✓	http://www.aajkaal.net
Abasar	✓	✓	http://www.abasar.net

Amader Malda	✓	✓	http://www.ndparking.com/amadermalda.com
Anandabazar Patrika	✓	✓	http://www.anandabazar.com
Bangla Ganashakti	✓	✓	http://bangla.ganashakti.co.in
Bartaman Patrika	✓	✓	http://bartamanpatrika.com
Century Sangbad	✓	✓	http://www.suprovat.com
Dainik Statesman	✓	✓	
Ek Din	✓	✓	
Gold News	✓	✓	http://www.goldnewsindia.com
Sambad Pratidin	✓	✓	http://www.sangbadpratidin.net
Uttar Banga Sambad	✓	✓	http://www.uttarbangasambad.com

Among the 12 dailies available, six of them have both print and on-line versions. These six are taken into consideration according to circulation and availability in both online and print forms.

HISTORY OF THE NEWSPAPERS

Aajkaal (আজকাল): Aajkaal is published from Kolkata. In 1981, the newspaper was started by Abhik Kumar Ghosh. The newspaper is known for left-leaning coverage and its comprehensive sports news. Its readership was measured at 33 lakhs by the Indian Readership Survey (IRS) 2008. The first editor of the newspaper was legendary journalist Sri Gour Kishore Ghosh. Last 30 days newspaper is available at online archives.

Anandabazar Patrika (আনন্দবাজার পত্রিকা): Anandabazar Patrika is published from Kolkata, New Delhi and Mumbai by ABP Pvt. Ltd. Anandabazar is the most circulated Bengali newspaper in India. The paper was founded in 1922 by its inaugural editor Prafulla Chandra Sarkar. Presently, the newspaper is edited by Aveek Sarkar. The circulation is more than a million a day. Other than Kolkata, it is also printed from various other towns in West Bengal.

It is also published from Guwahati (for entire northeast), Siliguri (for North Bengal and Sikkim), Jamshedpur, and Ranchi (for Jharkhand).

Online version was launched on 1st April 2001. The online archives hold all the previous issues.

Bartaman Patrika (বর্তমান পত্রিকা): Bartaman Patrika is published from Kolkata by Bartaman Pvt. Ltd. The newspaper was founded by Barun Sengupta, a journalist of Calcutta in December 1984. Apart from the Kolkata edition, the newspaper has four other simultaneous editions, published daily from Siliguri, Burdwan, Malda and Midnapore. The newspaper competes well with Anandabazar Patrika. From April 2002, Bartaman Patrika has published an online version.

Ganashakti (গণশক্তি): Ganashakti is the official mouthpiece of the Communist Party of India (Marxist) West Bengal State Committee. The paper first appeared as a fortnightly in 1967. After that, the newspaper was brought out as an evening daily for quite some time and finally converted into a full-fledged daily newspaper. Even as a party mouthpiece, it covers wide range of issues like,

literature, science, and technology. It is also published in Bengali and English weekly editions, and is read in the city of Kolkata and the surrounding area. A unique feature of the newspaper is its display on roadside woodboards throughout Kolkata and its suburbs allowing people who can't afford a newspaper to read it.

Since 1st September 2004, Ganashakti has published an online version.

Sangbad Pratidin (সংবাদ প্রতিদিন): Sangbad Pratidin is published from Kolkata and owned by Swapan Sadhan Basu. The paper started publishing from 9 August 1992. This was the first Bengali newspaper to start an online edition. Initially the newspaper maintained political neutrality but has become the virtual mouthpiece of Trinamool Congress as news and views published are always strongly biased towards Trinamool congress. Kanad Dasgupta is the present chief reporter of this newspaper.

From 2003, Sambad Patrika has published an e-paper. It is the first newspaper in Bengali who has its own website. The online archive holds the issues from June 2008.

Uttar Banga Sambad (উত্তর বঙ্গ সংবাদ): Uttar Banga Sambad is published from Siliguri. Uttar Banga Sambad was started on 19 May 1980 in a small letter press in Siliguri. With its huge popularity, in 1981 web offset press was installed. It is the largest circulating daily in North Bengal and have a share of about 80% out of all other dailies in the region. It is now printed simultaneously from Siliguri, Cooch Behar and Malda.

From 2003, Uttar Banga Sambad has published an online version. The online archive holds the issues from April 2008.

PRESERVATION METHODS

Different ways are adapted for preservation of newspapers. Traditional method of preservation is those of printed newspapers i.e., the hard copy. Keeping news clippings/press clippings were the only means in good old days for newspaper preservation. Different newspaper archives keep the news clippings in an organized way so that it can answer the queries of the users. Microfilm is an important method in newspaper preservation as huge amount of newspapers can be stored in a little microfilm roll and this remain intact for next 500 years. Digital preservation is the way to provide quick and easy access. E-paper archives are now available on the online versions of newspapers.

IN CASE OF HARDCOPY

Newspaper clippings / Press Clippings

Press clippings is an important form of preservation as to retrieve the required information. It is a process that involves cutting and pasting of important articles and their organization. It is the most important service in the newspaper library. Different newspapers follow different procedures for organizing newspaper

clippings (Gatos 2001). Anandabazar Patrika, Bartaman, Aajkal and other newspapers use clippings of important articles, index them and store them as a method of preservation.

Bound Volume / Binding

Ideal environment for preservation of newspapers is dark, 60-70°F and 40-50 per cent relative humidity. Back volumes of newspapers are generally bounded and kept together at 70°F temperature and 40% humidity by Anandabazar Patrika and Bartanan. At Pratidin previous issues are tied down month wise and stacked, pesticides are used at regular intervals.

Encapsulation

Encapsulation was developed by the Library of Congress to protect fragile and brittle newspapers. The process involves placing the sheet of newsprint between two clear sheets of polyester film. The film is sealed around the edge with a double stick tape or by machines designed to bond the film. The papers which are encapsulated or placed in plastic sleeves generate acidic gases and will deteriorate the paper. Deacidification is necessary for all the papers that are to be encapsulated (Das 2009). But, for all the Bengali newspapers as mentioned, no encapsulation is done as it is an expensive process.

Lamination

Newspapers are a continuous diary of our world. When an historic event occurs, when people find evidence to be used in genealogical research, and when there's a report about a personal accomplishment, people will want to clip the article and do their best to preserve the newsprint. Very old and partly damaged newspapers are kept after proper processing and lamination. By using high-quality, acid-free, alkaline-buffered materials, lamination should be done in proper way. As newspaper is quite big in size, lamination should be done very carefully and part by part. Among the Bengali newspapers lamination of important clippings are done by Anandabazar Patrika only.

FROM HARD COPY TO SOFTCOPY

Microform: film / fiche

Microform preservation of newspapers was first introduced by microfilming the *London Evening News*, in 1853. Life expectancy of polyester-based microfilm is 500 years (approx.). Most of the newspaper establishments have microfilms as it's the most important thing in newspaper preservation. Microfilms can be digitized by scanning. The scanned images are converted into digital TIFF format. But still the search with keyword and indexing is not possible. The Photo unit of Anandabazar Patrika also stores the microfilm collection. The microfilms are obtained from the Library of Congress Photo Duplication unit. In addition to their own publications, they have microfilm collections of some other newspapers such as *The Economist* and *The Times*. Their future plan is to go digitizing.

Digitising

A way of capturing and storing images using computer technology is called digitising – a digital camera or a scanner takes images of the document and converts it into binary code. Optical Character Recognition (OCR) can be used not only as editable text as well as for refinement of letters. In case of Bengali language there are no good quality OCRs available. Depending upon the quality of the original document the digitalized papers can be stored both in Hard Disk and back up in DVD/CD-ROM. Bangla Ganasakti is available online in the form of scanned images (digitized) of the print version. Section wise link is made to increase the size of the article; it therefore doesn't have proper indexing system and requires a lot of space to store the newspapers. Anandabazar Patrika, Bartaman Patrika and Pratidin have digital copies.

Digital Preservation

The advent of computer technology has added the new dimension to the publishing of newspaper. Besides the digitization of old microfilm, newspapers have come up with e-paper for Internet and for easy archiving. E-paper is the new demand of the users as it is available on the World Wide Web. Users can easily search, print, and email articles and advertisements (Jantz 2005).

Digital preservation is defined as the managed activities necessary for the long term maintenance of a byte stream (including metadata) sufficient to reproduce a suitable facsimile of the original document and for the continued accessibility of the document contents through time and changing technology. A digital repository is simply a "place" to store, access, and preserve digital objects. Digital objects can be quite complex reflecting the structure of the physical artifact and including multiple content byte streams and special software used to deliver dynamic results to the user (Jantz 2003). A flexible digital repository should allow storing all types of digital objects along with the appropriate descriptive and administrative metadata. A digital object might be an electronic journal article, a digitized image of a photograph, numeric data, a digital video, or a complete book in digital form. But unfortunately all the Bengali newspapers store their softcopies in a format which is quite different from that of their print versions except in case of Ganasshakti. The main reason is the non availability of good quality OCR. They keep their digital versions on hard disk as well as on DVDs. Ganasshakti stores the scanned image (digitized) of the print version.



PROBLEMS

Bengali newspapers have the following problems -

- Binding has been a frequently used method for organising and storing newspaper files for many decades, but it is not desirable. Binding of newspapers often damages the text; creates huge volumes that are difficult to handle properly.
- A space consideration is a problem in hardcopy repositories for all the agencies as the volume increases day by day. Solution(s) must be comprehensive to manage this problem, yet efficient and sustainable in the long term.
- Searching and indexing is very much difficult in hardcopy as it is a complex and some tire job. Users often face difficulties to find right information at right time.
- Image storing (TIFF) takes lot of hard disk space and doesn't have content based searching facility, only section or article based searching is possible (Lynch 2001).
- Bit stream is used in the website for the Bengali fonts by all the newspapers in their online versions which has to installed in the client server by the user. Only Internet Explorer is supported and not Firefox or Opera.
- UNICODE can be used instead of bit stream but proper designing of ISCII compliant Bengali True Type Font (TTF) and designing UNICODE compliant Bengali Open Type Font (OTF) is a challenging issue- it's still in a research phase.

CONCLUSION

Newspapers are considered to be the first drafts of history, while at the same time, are part of a country's cultural heritage. Preservation of newspapers an utmost necessity. Converting hardcopy archives to digital resources, digital preservation in terms of preventing paper deterioration as well as full utilization of the archives can be possible. Microfilm shows very low risk if proper film base, processing, and physical environment are maintained. Digital storage may have low to high risk as hardware and software change very frequently and life expectancy of the formats are still not known. Functionality of microfilm is uncompromised but when compared to digital storage, it provides more easy access to the desired information. Anandabazar Patrika has future plans for digitization of their back volumes with the help of OCR technology invented by Indian Statistical Institution (ISI), Kolkata. Finally, it can be concluded that financial support and technologically experienced persons are required for proper preservation of these Bengali newspapers.

REFERENCES

1. Das, S. 2009. Preservation of Newspapers. *DESIDOC Journal of Library & Information Technology*, vol. 29 (no. 1): 72-75
2. Gatos, B., Mantzaris, S.L. and Gouraros, N. 2001. An integrated system for creating a Digital Library from Newspaper Archives The Digital Resources for the Humanities (DRH2001) Conference, London, U.K., July 2001
3. Jantz, R. 2003. Public opinion polls and digital preservation: An application of the Fedora Digital Object Repository System. *D-Lib Magazine*, vol. 9 (no. 11) Available at: <doi:10.1045/november2003-jantz>. (accessed on 14th January, 2010)
4. Jantz, R. & Giarlo, M. J. 2005. Digital Preservation: Architecture and Technology for Trusted Digital Repositories *D-Lib Magazine*, vol. 11 (no. 6) Available at: < <http://www.dlib.org/dlib/june05/jantz/06jantz.html>>. (accessed on 14th January, 2010)
5. Lynch. C. 2001. When documents deceive: Trust and provenance as new factors for information retrieval in a tangled web. *Journal of the American Society for Information Science and Technology*, vol. 52 (no. 1): 12 - 17.
6. Wikipedia Newspaper (html) <http://en.wikipedia.org/wiki/Newspaper> (accessed on 17.12.2009)

GROWTH AND DEVELOPMENT OF ONLINE NEWSPAPERS WITH SPECIAL REFERENCE TO INDIA

Dr. Sumeer Gul

and

Tariq Ahmad Shah

ABSTRACT

Design/Methodology: *The paper examines the results from a survey conducted from well renowned online newspaper directories. The research covers the following aspects: global development of online newspapers, causes for growth, Indian perspective, coverage, language diversity, state and city output, archiving policy, and impact of web 2.0 on the online newspapers and their associated services.*

Findings: *5786 online newspapers host the web from 178 nations. USA ranks first in the list while India occupies the 4th position. The results show a mounting drift of online newspapers especially among the developed nations. India, which is also showing an unremitting growth towards online newspapers, has yet to compete with the rest of the world. Overall, a compound annual growth rate (CAGR) of 28.29 per cent is observed over a period of 14 years (1996 to 2009). Maharashtra leads among the 31 contributing states/union territories of India. Indian newspapers are available in 21 languages with a high percentage (92.69%) with monolingual interface and about 36 per cent in English and above 16 per cent in Hindi. Language barrier, a biggest impediment in communication has yet to be solved and research has to be carried out to provide a multilingual access to the online content through newspapers. E-paper facility is found in around 58 per cent while approximately 60 per cent provide archive facility. A meager proportion of about 15 percent provide video links. Around 30 per cent make use of web 2.0 tools. Web 2.0 which is the latest buzz and holds a paramount importance in the online market has yet to make its way in the Indian online newspaper souk.*

Future Implications: *An intense study on user behaviour and online newspapers can be carried out using deep log analysis methods. Studying the needs of the online newspaper patrons in a virtual world will help in designing a better sustainability for online newspapers.*

Originality/Value: *The study presents the results from the data gathered from different tools which provide access to online newspapers worldwide and India particularly. The study depicts originality in its contents and is of value especially to the reference librarians, information professionals, media personnel and others interested in grabbing nascent information through newspapers.*

Keywords: Online newspapers, Online Journalism, Online Periodicals, Web 2.0, Online newspapers-India, Reference Services, India.

INTRODUCTION

It is a well known proverb, 'one can do without a cup of tea in the morning, but not without a morning newspaper'. But a number of obstacles hinder the people to have access to information, they are in need of. The World Wide Web has opened new dimensions for the information flow with the advent of web resources which have become a rivulet crossing all the borders (Gul 2007). Online newspapers, a promising medium for dissemination of information are increasing their magnitude on the Web. The potentialities of the Web have been accepted by the newspaper market globally. The shrinking market of the traditional print newspapers led the print news to drift to the World Wide Web and its dynamic delivery realized a need to have web counterparts of old traditional newspapers. The newspapers available via the Web are making a mainstream medium that play as important a role as their printed counterparts in delivering information and informing the public (Eveland et al 2004). Their rise has been one of the most spectacular success stories of the Internet as has been witnessed by Li. He furthermore comments on online newspapers as a more mature, more sophisticated, and more efficient medium in delivering information (2006: ix). The newspapers on the Internet were born with a much lower profile, though not free from the prophecy of media analysts (Riley et al 1998). Not as many as great expectations were placed on them compared to other new media derived from technological innovations, (Giles 2000; & Ingle 1995:17) online newspapers not only survived, but also expanded exponentially. Online newspaper is a phenomenon along with the growth of Internet.

The growth of online newspapers traces back to 1992 when for the first time U.S newspapers were published on Web. Newspaper content was text based and was delivered through BBS or online services such as Prodigy and America Online. On September 12, 1994, Netscape released the beta version of its navigator, a graphic web browser. Newspapers formally established their presence on the World Wide Web soon after that. There were approximately 60 North American newspapers with sites on the Internet or with dial up services by the end of 1994. The pioneers among the online newspapers were Raleigh News and Observer, San Francisco Observer/ Chronicle, and San Jose Mercury News. In 1996, approximately 500 North American newspapers established sites on Internet or through dial up services, but only half of them, 248 daily newspapers were published on the Web by September 1996. Newspapers on the Web reached the critical mass scale around 1997. By September 1997, 745 U.S newspapers, about half of the U.S dailies were published on the Internet as witnessed by Meyer (cited by Li 2006:2). The growth of newspapers on the Internet accelerated in 1997. There were 1,290 U.S newspapers online in March 1998. The number reached 2,059 (as of September 27, 1998), a growth of 60% in six months and 176% in one year (Peng et al 1999:52-63). More than 3,400 U.S papers were online by June 2001. Most of the U.S newspapers were published on the Internet in 2001 as witnessed by Poynter (cited by Li 2006: 2). The unprecedented magnitude of online newspapers is expanding gradually.

Many factors are responsible for their popularity. On many occasions, audiences rely more on newspapers on Internet for information because of ease in access,

updated more often, and richer in content than print newspapers (Chyi and Lasorsa 1999: 2). Publishers can use the Internet as a marketing channel to reach potential customers that do not read print newspapers. They can also use the Internet to strengthen relationships with existing readers by offering new services (breaking news, classified and sites), by collecting reader information, and by establishing two-way communication with readers (Mings & White 2000; Peng, et al 1999:52-63) (cited by Wurff and Lauf 2005:14). Online editions help in making the content available in real-time, and text would be sexed up with moving images and sound (Boczkowski, 2002; & Massey & Levy, (1999) (cited by Wurff and Lauf 2005:15) which is the choice of today's world.

Online newspapers are becoming the present trend in the virtual world with extraordinary intensity. Newspapers, being a product of printing press enjoyed the privilege of monopolizing the mass media market for centuries until the advent of radio and television. It was World Wide Web that helped the newspaper sector to flourish (Peng et al 1999:52-63). Time is not far when newspapers in a virtual mode will become the choice of globe.

REVIEW OF LITERATURE

Although the concept of online newspapers is not a new one, but very less research has been carried out regarding their growth and development. The technological development has brought about unparalleled and drastic changes in the newspaper market. The concept of online newspapers has opened new horizons for the seekers of nascent thoughts. The year 1995 saw newspapers settling on the Web as their electronic publishing environment of choice (Beamish 1997; Carveth, Owers & Alexander 1998; Garrison, 1997; Martin and Hansen 1998; Molina, 1997) (cited by Boczkowski 2002: 273). An early study of online newspapers found that interactive features in them were scarce as opined by Tankard and Ban (1998). A count by the Newspaper Association of America showed more than 175 US dailies publishing on the web at the end of 1998, a number that grew to over 750 three years later (Newspaper Association of America, 1998; Editor and Publisher, 1996) (cited by Boczkowski 2002: 273). Chyi and Lasorsa (1999:2) studying the access, use and preference for online newspapers witnessed a total of 2,900 newspapers operating online, of which more than 1,800 were U.S based. Li (2006: 2) researched that from the period of 1996-2002; more than 4,000 newspapers were created on the Internet.

OBJECTIVES

The study aims to explore the online newspaper wealth in general at global level and India in particular. The study delves deep in to the coverage, language diversity, state output, archiving policy, and impact of web 2.0 on the online newspapers and their associated services.

SCOPE

The study has taken into account only those Indian Online newspapers which are also published in print, and the newspapers having only online internet edition were omitted.

METHODOLOGY

To build a picture of online newspapers worldwide, papers listed in two excellent and comprehensive directories, Ulrichsweb.com and Internet Public library (IPL) were clubbed together. As each noted different figures for the same country, duplicate entries were omitted in order to arrive at an appropriate figure. A number of websites like Yahoo Newspaper Directory, Paperboy.com, Online Newspaper.com, and Newspaper Association of America etc, which list online newspapers, were also consulted. However, on manually checking for Indian perspective, a number of errors were traced out and as such were excluded for global view. Since the main thrust of this study is to gauge Indian output, thorough and exhaustive search was conducted on a number of websites besides the above mentioned websites. The entire run of each newspaper was examined manually to authenticate their existence and to harvest the necessary data. Besides, Registrar of Newspapers for India (RNI), an apex body to compile and maintain register of newspapers published (in print) in India was also consulted in order to ascertain their print mode. In addition, database of Whois (<http://who.is/>), a web domain information provider, was consulted to retrieve information regarding the year of launching online edition and contact information of the newspaper website holder.

RESULTS

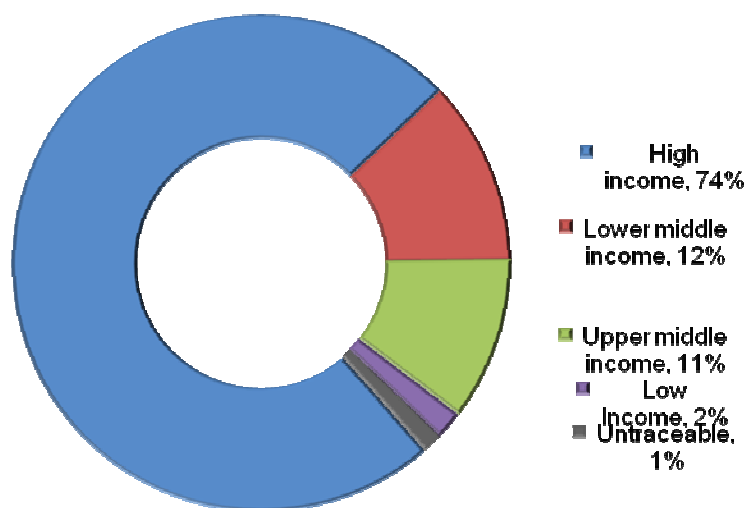
As of 24 December 2009, Registrar of Newspapers for India (RNI) recorded an approximation of 7532 distinct titles that are daily published in around 41 languages from 7532 cities of 35 Indian states/union territories. However a meager number of 260 papers is found to have presence on the web.

Global Perspective

A total of 5,786 online newspapers from 178 contributing countries are found to enrich the Web. USA proved to be a giant in terms of number of online newspapers (2402), accounting for 41.51 per cent of the total (Table 1). Canada and United Kingdom follows the list with 327 (5.65%) and 306 (5.29%) newspapers respectively. India holds 4th position with a total of 260 (4.49%) online newspapers. Moreover, the World Bank List of Economies List (July 2009) was applied which categorized the nations in to High Income; Upper Middle Income, Lower Middle Income and Low Income Countries. The study clearly reveals that 74% of the total online newspaper wealth is from High Income countries followed by Lower Middle Income (12%) and Upper Middle Income (11%) countries respectively. Low Income countries accounted a meager percentage of 2% only. (Fig. 1)

Table 1: Country wise distribution of Newspapers

Rank	Country	No. of Newspapers	Percentage
1	United States of America	2402	41.51
2	Canada	327	5.65
3	United Kingdom	306	5.29
4	India	260	4.49
5	Germany	232	4.01
6	China	139	2.40
7	Mexico	136	2.35
8	Russia	99	1.71
9	Italy	96	1.66
10	Spain	85	1.47
11	Australia	83	1.43
12	Norway	75	1.30
13	France	61	1.05
14	Denmark	54	0.93
47 countries with newspapers 10-50		1023	
26 countries with newspapers 5-10		188	
37 countries with newspapers 3-5		148	
18 countries with 2 newspaper each		36	
36 countries with one newspaper only		36	
TOTAL		5786	

**Figure 1:** Distribution of newspapers on GDP status of countries**State wise distribution**

State wise distribution which is based on the location of main office/headquarters of a newspaper reveals that Maharashtra with 34 newspapers or 13.08 per cent of total ranks 1st among the 31 contributing states/union territories of India. Delhi with 24 papers, accounting to 9.23% follows the list.

Gujarat and Karnataka with 19 papers each (7.31%) held the third rank. As indicated in table 2, among 5 top ranked states, 4 positions are occupied by those states having metropolitan cities (Delhi, Kolkata, Mumbai, & Chennai). Ironically, no online paper was found from Bihar during the study period. (Table 2)

Table 2: State wise distribution of Online Newspapers

Rank	State	Newspapers
1	Maharashtra	34 (13.08)
2	Delhi	24 (9.23)
3	Gujarat	19 (7.31)
3	Karnataka	19 (7.31)
4	West Bengal	18 (6.92)
5	Tamil Nadu	17 (6.54)
6	Andhra Pradesh	15 (5.77)
7	Jammu and Kashmir	13 (5)
7	Kerala	13 (5)
8	Assam	12 (4.62)
9	Madhya Pradesh	9 (3.46)
9	Uttar Pradesh	9 (3.46)
10	Orissa	7 (2.69)
11	Punjab	6 (2.31)
11	Rajasthan	6 (2.31)
12	Arunachal Pradesh	5 (1.92)
12	Nagaland	5 (1.92)
13	Goa	4 (1.54)
14	Chhattisgarh	3 (1.15)
14	Manipur	3 (1.15)
14	Sikkim	3 (1.15)
15	Chandigarh	2 (0.77)
15	Haryana	2 (0.77)
15	Jharkhand	2 (0.77)
15	Meghalaya	2 (0.77)
15	Tripura	2 (0.77)
15	Uttarakhand	2 (0.77)
16	Andaman Nicobar	1 (0.38)
16	Himachal Pradesh	1 (0.38)
16	Mizoram	1 (0.38)
16	Pondicherry	1 (0.38)
Total		260 (100)

Figures in parentheses indicate percentage

Language

A high percentage of papers (241, 92.69%) are monolingual succeeded by 17 papers (6.54%) as bilingual and a meager number of 2, with multilingual interface. As shown in Table 3, Indian newspapers are available in 21 languages with highest number of 102 papers (36.3%) in English followed respectively by Hindi (16.37%) and Gujarati (6.76%).

Table 3: Language wise distribution

Rank	Language	No. of Newspapers
1	English	102 (36.3)
2	Hindi	46 (16.37)
3	Gujarati	19 (6.76)
4	Urdu	15 (5.34)
5	Marathi	14 (4.98)
6	Kannada	13 (4.63)
7	Malayalam	12 (4.27)
8	Tamil	10 (3.56)
9	Assami	9 (3.2)
9	Bengali	9 (3.2)
9	Telgu	9 (3.2)
10	Oriya	7 (2.49)
11	Punjabi	6 (2.14)
12	Manipuri	2 (0.71)
12	Nepali	2 (0.71)
13	AO	1 (0.36)
13	Goan Konkani	1 (0.36)
13	Khasi	1 (0.36)
13	Mizo	1 (0.36)
13	Sanskrit	1 (0.36)
13	Tibetan	1 (0.36)

Figures in parentheses indicate percentage

Chronological Development

Table 4 visualizes that Indian newspapers started to host their content on web from the year 1996 and a big leap of 328.57% was observed in the year 1997. The year 2007 witnessed maximum number of newspapers (32) that flashed their content online followed by the year 1998 and 2005 with 27 and 25 newspapers respectively. Overall, a compound annual growth rate (CAGR) of 28.29% is observed over a period of 14 years (1996 to 2009). (Fig 2)

Table 4: Distribution of newspapers over the year

Year	No. of Papers	Cumulative No.	Cumulative Growth
1996	7	7	-
1997	23	30	328.57
1998	27	57	90.00
1999	22	79	38.60
2000	14	93	17.72
2001	9	102	9.68
2002	13	115	12.75
2003	7	122	6.09
2004	10	132	8.20
2005	25	157	18.94
2006	18	175	11.46
2007	32	207	18.29
2008	22	229	10.63
2009	17	246	7.42
UT	14	260	-

UT Untraceable

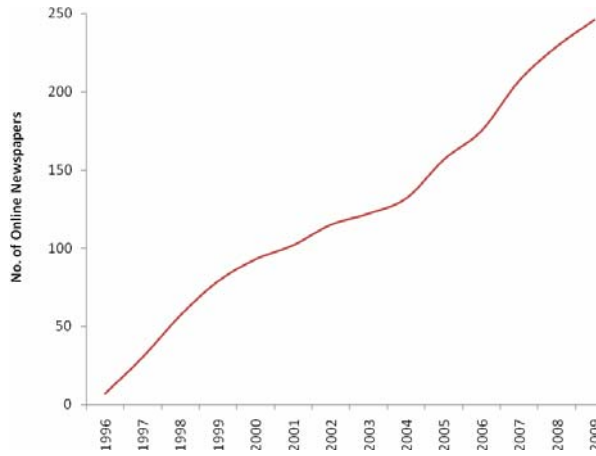


Figure 2: Cumulative Growth over a span of 14 years (1996-2009)

e-paper & Digital Preservation Policies (Archive)

e-paper, a technological advancement to read and experience printed newspapers online, was available for 152 newspapers amounting to 58.46%. Newspaper sites which provide access to the back issues at least for one week were categorized under those papers with archive facility and as such a total of 155 papers provided access to the electronic archives. Besides, 41.54% (108) provide both facilities while 62 papers (23.85%) possess none. Table 5 supported with Fig. 3 gives a more clear view of the above findings.

Table 5: e-paper & Archive Facilities

	e-paper	Archive
Yes	152 (58.46)	155 (59.62)
No	108 (41.54)	105 (40.38)

Figures in parentheses indicate percentage

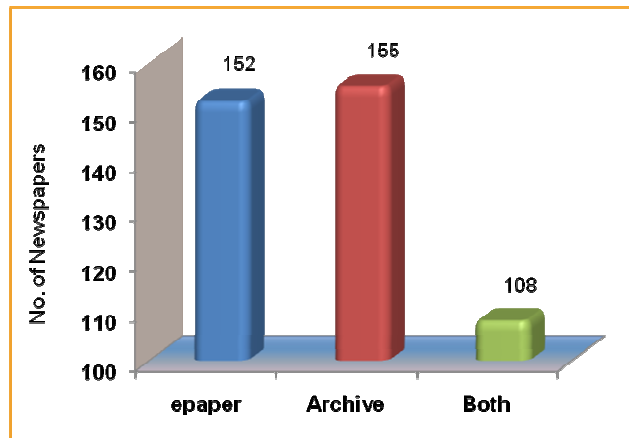


Figure 3: e-paper & Archive Facilities

Video Facility

Papers with their own video library and those providing links to the You Tube were tagged in the video category. From Table 6, it is clear that 40 (15.38%) papers provide video facility and an outsized proportion lacks this mesmerizing feature.

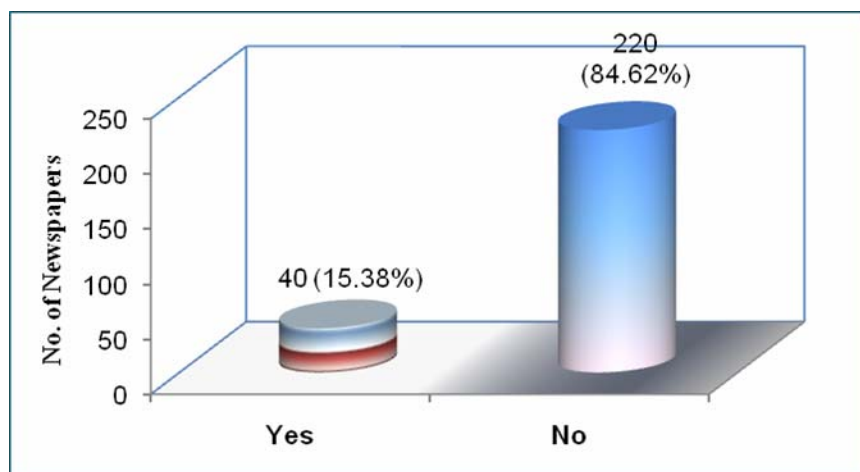


Figure 4: Video facility

Web 2.0 features

The study clearly reveals that three Web 2.0 tools are mostly used by the online news media, exemplified by Really Simple Syndication (RSS), Twitter, and Blog. Among these, RSS showed presence in 71 newspapers followed respectively by blogs in 37 papers (Table 7). Twitter being the recent invention of 2006 appeared least (37, 14.23%). In general, 78 newspaper websites utilize any of these features while a high percentage (70%) lacks Web 2.0 tools. Fig. 5 conveys that only 11 papers are having all these features, 21 with RSS and blogs, 6 possess RSS and Twitter and only one provides blog and Twitter.

Table 7: Web 2.0 features

	RSS	Twitter	Blog	Any
Yes	71 (27.31)	21 (8.08)	37 (14.23)	78 (30)
No	189 (72.69)	239 (91.92)	223 (85.77)	182 (70)

Figures in parentheses indicate percentage

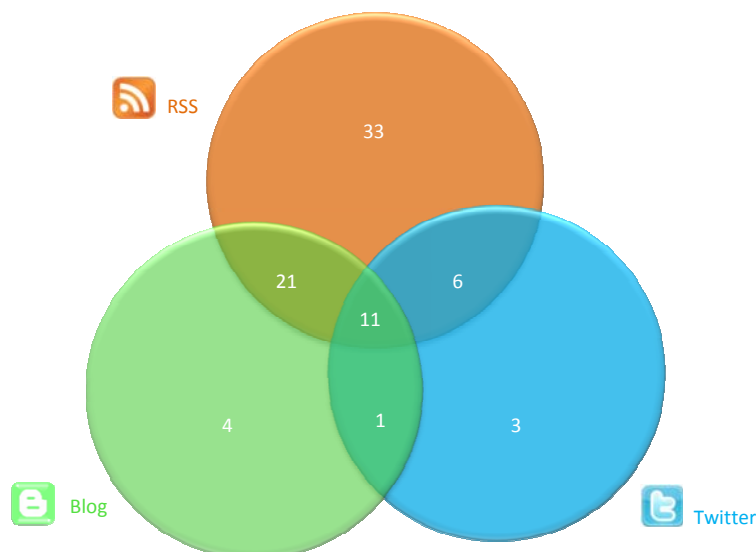


Figure 5: Venn representation of Web 2.0 features of Newspapers

CONCLUSION AND DISCUSSION

Online newspapers - 'the information highway' have established themselves as a viable and promising publishing media. The study clearly reveals that there is a flurry of growth of the online newspaper wealth globally as well as in India. The main reason for their growth at a global level is directly proportional to the high rate of growth in G.D.P which makes the High Income countries to give a better output in terms of online newspaper wealth. However, a developing economy like India is doing excellently in the online newspaper market because India has also followed the voice, 'information for all'. In addition, India has jumped on to Big Emerging Market (BEM) league. Furthermore, countries having high Human Development Index (HDI) also excel in online newspaper market because a promising HDI takes into account how income is turned into better opportunities. In terms of the state wise growth of online newspapers, central government has to develop a well knitted network with the state governments to evolve up with a promising number of online newspaper wealth from the local regions. They will not only be the best sources of information on the local current events but will also help in exploring the way outs for the problems that will be highlighted through the virtual mode of Tim Berners Lee, the mastermind behind Web. The technological impact has to be accepted by the states which have not yet tested waters with it in order to compete in this bit and byte world which believes in access rather than ownership. Information access, which has many a times been hindered because of the language problem, has to be researched a lot in a country like India where diversity in languages changes with every changing region. Newspaper companies have a vast field to explore in terms of language diversity to come up to the expectations of the users. If they will be in a position to provide access to their newspaper content in a multilingual mode, that can solve one of the biggest problems of the news

seekers. Not only regional languages should be preferred but languages of international origin will also help in developing better access opportunities. Collaborations with search engines like Google and Alta Vista that provide translation service should be taken care of by the newspapers agencies to eradicate the language barrier and to achieve a smooth communication.

The chronological growth indicates that the online sun is going to dominate the newspaper industry in a virtual mode soon.

However, to build future it is very important to preserve past because future stand on the shoulders of past. Digital preservation, which is an answer to this, can help in combating the problem of preserving the past. Archives should be maintained that can ensure preservation. Mirror sites should also be developed by the newspaper industry in order to ensure long term digital preservation.

Users which should be the first priority of the newspapers industry should be taken care of. Their needs should be catered in every possible way. In today's world where we talk of e-content, e-paper policy should be adopted by every newspaper that offers virtual access to the information.

Multimedia facilities like videos should also incorporated as multimedia has a tendency to attract people especially the younger lot.

Web 2.0, which is a buzzing phenomenon in the digital world and has taken hold, with more than 9.5 million citations in Google, (definition posted on the Web on September 2005) 135 million citations as of February 2007 (O'Reilly 2007: 17) has not left a trace of it on majority of Indian online newspapers. However, a meager number of online newspapers are adhering to this interactive technology which is having a tendency to deliver rich user experiences. Where, the world is moving towards Web 3 (also referred to as Semantic Web), a number of Web based Indian newspapers have yet to evolve drastically with this promising technology, i.e. Web 2.0.

The study clearly reveals that the future of online newspapers is bright not only in the developed nations but also in developing nations like India also, provided the Indian newspaper industry will take the online revolution in a positive manner and this medium should be made more interesting by adding more value added services that can act as an appetizer for the users.

REFERENCES

1. Boczkowski, Pablo J. 2002 The Development and Use of Online Newspapers: what research tells us and what we might want to know. *In* The Handbook of New Media: social shaping and consequences of ICTs. Leah A. Lievrouw and Sonia M. Livingstone, ed. New Delhi: Sage Publications
2. Chyi, H. I, and Lasorsa, D. 1999 Access, use and preferences for Online Newspapers. *Newspaper Research Journal*, 20(4). <http://www.questia.com/PM.qst?a=o&d=5002320220>, accessed December 12, 2009

3. Eveland, W. P, Marton, K., and Seo, M 2004 Moving beyond “Just the Facts”: The Influence of Online News on the Content and Structure of Public Affairs Knowledge. *Communication Research*. 31 (1): 82-108. <http://crx.sagepub.com/cgi/reprint/31/1/82>, accessed December 23, 2009
4. Giles, Bob. 2000 Journalism in the Era of the Web. *Nieman Reports*. Accessed February 2, 2010 from <http://www.highbeam.com/doc/1G1-71190366.html>
5. Gul, Sumeer. Web Resources: Online Journals, Online Books and Electronic Theses and Dissertations - A Wave of the Future. *International Journal for Technical Communication (IJTC)*, 2 (1)
6. Ingle, Bob. 1995 Newspapers vs. online versions: A discussion of the old and the new media. *Nieman Report*, 49 (2).
7. Li, X, ed. 2006 *Internet newspapers: the making of a mainstream medium*. New Jersey: Lawrence Erlbaum Associates.
8. O’Reilly, Tim. 2007 What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & strategies*, 65(1)
9. Peng, Foo Yeuh, Tham, Naphtali Irene, and Xiaoming, Hao. 1999 Trends in Online Newspapers: A Look at the US Web. *Newspaper Research Journal*, 20 (2). <http://www.questia.com/PM.qst?a=o&d=5002320220>, accessed January 12, 2010
10. Riley, P, Keough, C.M, Christiansen, T, Meilich, O., and Pierson, J. 1998 Community or colony: The case of online newspapers and the Web. *Journal of Computer-Mediated Communication*, 4(1). <http://jcmc.indiana.edu/vol4/issue1/keough.html>, accessed January 01, 2010
11. Tankard J. W. and Ban, H.1998 Online newspapers: Living up to their potential? *Paper presented at the 81th annual meeting of the Association for Education in Journalism and Mass Communication*, Baltimore.
12. Wurff, R. V. D, and Lauf, Edmund, ed. 2005 *Print and online newspapers in Europe: A comparative analysis in 16 countries*. Amsterdam: Het Spinhuis publishers

CONCEPT ANALYSIS OF NEWS FOR MINING STRATEGIC INFORMATION

Sumit Goswami

and

V Senthil

ABSTRACT

A concept map provides visualization of the news events held over a period of time by extracting the main concepts from the news dataset and how they are interlinked. A concept is a set of weighted words that generally travel together throughout the text. This paper processes and displays the concept structure of all science, technology and defence related news events reported in major newspapers of India during 2008. The news dataset passes through various stages namely identification of news dataset, preprocessing, concept identification, concept editing, thesaurus learning, locating concept and finally concept mapping so as to make it suitable for effective visualization. The output of the process displays the main concepts in news dataset, identifies proper names i.e. named entity recognition (NER) and generates a thesaurus from the news events. Similar concepts are clustered together into themes. We also plan to integrate a summarizer along with the concept map so as to provide a summary of a selected news event from a concept map. The concept analysis is done by measuring the presence and frequency of a concept which can be explicit like words, phrases etc or implicit like concepts itself. So as to identify the starting point for defining the concepts, a concept seed word is extracted. During the process of concept learning, a thesaurus is generated as a spin off of the process. The various issues related to concept analysis and the appropriate levels of analysis are also discussed in the paper.

Key Words: Conceptual Analysis, Content Analysis, Concept learning, news visualization, text mining

CONTENT ANALYSIS

Content analysis is the process of determining the existence of a set of words which may form a concept in a set of unstructured data like textual document, blogs, emails etc. The analysis of contents supports identification of the text into various known categories, which can be further used to establish a relationship among each other. An effective analysis of content can lead to proper understanding of the text and getting an insight into it. It may support stylometric analysis of the text as well as authorship identification, demographic detection, time stamping of the material and other quantification of the text data. Content analysis involves a critical research path as it can lead to identification of any communication form. Thus it can be used, for example, by social

psychologists to compare groups of individuals non-invasively by analyzing their natural social interactions without the need for creating artificial settings or scenarios. Furthermore, as news documents tend to exist over long periods of time, the technique can be used to extract valuable historical and cultural insights.

As content analysis can be performed on numerous forms of data ranging from political speeches and open-ended interviews to newspaper articles and historical documents, it is invaluable to many researchers. Such uses include:

- historical analysis of political speeches
- detecting the existence and level of propaganda
- coding surveys that ask open-ended questions
- determining the psychological state of the writers
- assesses textual content against measures
- assess cultural differences in populations

Types of Content Analysis

In general, approaches to content analysis fall into two major categories: conceptual analysis and relational analysis. In conceptual analysis, news documents are measured for the presence and frequency of concepts. Such concepts can be words or phrases, or more complex definitions, such as collections of words representing each concept. Relational analysis, by contrast, measures how such identified concepts are related to each other within the documents.

Conceptual Analysis

Conceptual analysis, also known as thematic analysis is the most common form of content analysis. Conceptual analysis involves the detection and quantification of the presence of predefined concepts within the text. Concepts can either be explicit i.e. particular words or phrases or implicit i.e. implied, but not explicitly stated in a set of predefined terms. As expected, explicit concepts are the easiest to define and measure, with the quantification of implicit concepts requiring the use of specialized dictionaries or contextual translation rules.

Issues in Conceptual Analysis

Although conceptual analysis may seem pretty straightforward, there are many decisions that need to be made by the researcher, each of which highlights important issues about the approach. Some of the main issues and how they had been dealt by us are discussed below.

What is the appropriate level of analysis?

Conceptual analysis is aimed at quantifying the presence of concepts in a given text. The first main issue raised is what exactly constitutes a concept? Coding may be attempted at many levels, including the detection of particular words or phrases, the coding of the sense of the word taking into account synonyms or

idioms, or the categorization of sentences, paragraphs or whole bodies of text. One popular approach to defining concepts is in the creation and application of general dictionaries. Such dictionaries explicitly define the possible coding categories, providing rules for assigning words to categories or providing the actual words to code for. Although the creation of general dictionaries minimizes the need for dictionary construction and validation, they are problematic in that they are specifically tailored to particular domains.

Assumed versus Inferred Categories

One problem in using general dictionaries for conceptual analysis is that they are generally useless in coding for documents outside their domain. Also, it is often argued that assumed category schemes impose the reality of the investigator on the text rather than measuring the categories used by the writers of the text themselves. That is, using different dictionaries, multiple descriptions of the same textural reality can be extracted with vastly different results, rather than the actual description meant by the authors. It is often argued that the categories present in the text can be inferred from the covariation amongst the high-frequency words in the text. One of the spin off of our content analysis is that it can automatically extract its own dictionary of terms for each document set using this information. That is, it is capable of inferring the concept classes that are contained within the text, explicitly extracting a thesaurus of terms for each concept. Furthermore, this process can be externally triggered with knowledge of the domain to influence the concept definitions, focusing on the concepts that we are interested in quantifying.

How do you distinguish concepts?

One decision that commonly needs to be addressed in conceptual analysis is whether or not categorical judgments are mutually exclusive. This is a problem because many statistical analysis procedures require variables that are not confounded i.e. correlated. However, schemes that are capable of coding for the meaning of words in the sentences rather than coding for particular words are flexible by necessity, often leading to ambiguities in categorical judgments. This problem is confounded by the fact that having one attribute does not necessarily exclude the possession of another. Furthermore, in many coding schemes, the qualities by which concepts are judged are continuous, reducing the reliability of category membership decisions made by human coders. To deal with the issue of continuous attributes in concept definitions, we weight each word depending of how well it connotes the category. In this approach, sentences may contain multiple concepts, but with the sentences needing to contain enough of the crucial keywords for a concept to be identified. The weighting of each word in the definition of a concept is automatically extracted.

Reliability

In content analysis, generally there are considered to be two forms of reliability that are pertinent: stability and reproducibility. Stability generally refers to the tendency of a coder to consistently re-code the same information in the same way over time. However, in using human coders, there are often inconsistencies that arise due to various factors such as the ambiguity of the coding rules,

ambiguities in the text, simple coding errors, or cognitive changes within the coder such as the presence of a good essay affecting the rating of subsequent ones. As our approach is automated and deterministic, many such inconsistencies are avoided, leading to a high level of coding stability. The second form of reliability, reproducibility, refers to the consistency of classification given by several coders, given the same marking scheme. In our case, this issue is most relevant to the generation of the conceptual map. That is, the process of map generation is stochastic, leading to the possibility of different final positions for the extracted concepts each time the map is generated. However, although the process is stochastic, there typically exist consistent trends in the spatial positioning of concepts relative to each other. To promote a high level of reliability, the map should be constructed several times, and interpretations of the data should focus on stable features.

Relational Analysis

Relational analysis, also known as semantic analysis goes one step beyond conceptual analysis, measuring the relationships between the identified concepts. Approaches to relational analysis generally fall within three categories: affect extraction, proximity analysis and cognitive mapping. Affect extraction, which aims at providing an emotional evaluation of the emotional and psychological state of the speaker, is a quite specific form of content analysis. The second form of relational analysis, proximity analysis, measures the co-occurrence of concepts found within the text. In this approach, a length of words or sentences called a window is specified. The window is moved sequentially through the text, with concepts that occur together within the window being noted. The result of this calculation is called a concept co-occurrence matrix, in which the frequency of co-occurrence of all concepts against all others is explicitly stored. This matrix gives a convenient metric for comparing the content of different document sets. The third form of relational analysis, Cognitive Mapping, is an extension of the first two, representing the information visually for comparison. As the output of relational analysis is generally a large matrix of values, the cognitive mapping approach tries to compress this information into two dimensions, displaying the main relationships between concepts.

CONCEPT

Concepts are collections of words that generally travel together throughout the text. For example, a concept *rifle* may contain the keywords *rifle*, *telescopic*, *serial*, *ammunition*, *stock*, *western cartridge*, etc. These terms are weighted according to how frequently they occur in sentences containing the concept compared to how frequently they occur elsewhere. Sentences are tagged as containing a concept if enough accumulated evidence is found. Terms are weighted so the presence of each word in a sentence provides an appropriate contribution to the accumulated evidence for the presence of a concept. That is, a sentence or group of sentences is only tagged as containing a concept if the accumulated evidence i.e. the sum of the weights of the keywords found is above a set threshold. For example, the concept *violence* may contain central keywords such as *war* and *terrorism* that provide strong evidence for the

concept in a sentence, but also may contain a collection of peripheral items such as *army* or *troops* that provide weaker evidence. Apart from detecting the overall presence of a concept in the text the concept definitions are also used to determine the frequency of co-occurrence between concepts. This co-occurrence measure is what is used to generate the concept map.

Concept Seed Words

Concept seed words represent the starting point for the definition of such concepts, with each concept definition containing one or more such seeds. These seed words can either be provided externally, or can be automatically extracted from the text. They are called seeds as they represent the starting point of the concept, with more terms being added to the definition through learning. Occasionally, more appropriate central terms may be discovered, pushing the seeds away from the centre of the concept definition.

Concepts and the Relevancy Metric

Concepts are defined as collections of terms that provide evidence for the use of the concept in the text. The terms themselves have a *relevancy value* that determines how central each word is to the concept. For example, in the Figure 1, *navy and war* will have a high relevancy value to *defence* as it appears often in sentences containing the concept, and rarely in sentences not containing the concept. In contrast, *technology* will have a low relevancy to *defence* as it co-occurs rarely with the concept, but frequently in the rest of the text. The concept seeds is identified by looking for candidates through the most frequently appearing words in the text that are not *stop words*. The potential seeds are next evaluated by calculating the number of strongly relevant terms for each seed candidate. This measurement selects important and central concepts that characterise the text. For example, the word *dog* may be considered a seed as it has many strongly related items such as *barks, fleas* etc., that often co-occur with it and occur not so often alone.



Figure 1 : Formation of concepts from relevancy

Concept Learning

During learning, a thesaurus of terms is generated for each concept. This learning is an iterative process in which the collection of terms defining a concept is updated. The aim of concept learning is to discover clusters of words which, when taken together as a concept, maximise the relevancy values of all the other words in the document. For example, the concept definition (*dog, hound, puppy*) is a strong concept definition as a large collection of other words (such as *fleas, bites, barks*, etc) frequently co-occurs with at least one of the concept keywords, and not so frequently elsewhere. This concept definition is considered stronger than a concept containing just *dog*, as the co-occurring words (*fleas, bites* and *barks*) may also appear frequently in other contexts (i.e. close to *hound* and *puppy*). The following are the steps of concept learning:

- Given the seed word (s), the relevancies of all other words in the document are calculated i.e. how often they co-occur with the seed item as opposed to how often they appear without it. For example if the initial seed word of a concept is *dog* the relevancies of the words *fleas* and *bites* might be high because they appear relatively frequently in blocks of text containing the concept i.e. containing the word *dog*
- Words are added to the concept definition if their relevancies fall above a certain threshold. For example if the concept seed is *dog*, the words *fleas, bites* and *barks* may be added to the definition due to their frequency of co-occurrence, leading to the new concept definition (*dog, fleas, bites, barks*).
- The process then continues, calculating the relevancy of other words in the document compared to the new concept definition i.e., how often does a word co-occur with any element in the concept definition as opposed to without any of them. These relevancy values are normalised, with the words above a set threshold being added to the definition. For example, given the concept definition (*dog, fleas, bites, barks*), the items *hound* and *puppy* might now be considered highly relevant due to their frequent co-occurrence with *fleas, bites* and *barks*.
- Due to their high relevancies, the words *hound* and *puppy* will be included in the new concept definition, leading to the concept (*dog, hound, puppy, fleas, bites, barks*).
- As the relevancies of the words contained within the concept are normalised and there is an inclusion threshold, over time certain of the initial keywords may be lost. For example, the words *fleas* and *bites* although initially quite relevant, may not be as relevant as the new keywords *hound* and *puppy* because they appear in other contexts. Thus, these words may be removed from the concept, leading to the concept (*dog, hound, puppy, barks*).
- Apart from adding highly relevant words to a concept, words that are negatively correlated with the concept i.e. words that rarely appear in sentence blocks containing the concept and frequently appear elsewhere may also be added. This was shown to improve the stability

of the convergence dramatically, with very common words which were not relevant contributing against classification.

- The process of learning is iterative, but will converge to a stable state. The learning halts when the number of sentence blocks classified by each concept remains stable.

PREPROCESSING

Preprocessing is the first phase of processing the news. This phase converts the raw documents into a useful format for processing. Preprocessing involves the following steps:

- ***Splitting the information into sentences, paragraphs and documents.*** These boundaries are important as they generally mark transitions in meaning. The conceptual map of the documents reflects the co-occurrence of distinct concepts. To prevent concepts from being perceived to be related across changes in context such as across different documents, the co-occurrence is only measured within and not across blocks typically containing three sentences.
- ***Removal of non-lexical and weak semantic information.*** Within each sentence, the punctuation is removed along with a collection of frequently occurring words called the stop-list that holds weak semantic information such as the words 'and' and 'of'. Furthermore, for documents extracted from Internet news groups, the headers are cleaned up and the non-text attachments are removed.
- ***Identifying proper names, including multi-word names.*** Often in documents the proper names such as people, places or company names depict important entities that should be mapped. For this reason, proper names are extracted as potential concepts. The words are classified as proper names if they start with a capital letter. As every word that starts a sentence falls into this definition, only start-of-sentence words that are not in the predefined stop-list are considered as names.
- ***Optional prose test of each sentence.*** To remove non-textual material from the text, such as menus and forms in web pages, sentences that are unlikely to be part of the specified language are removed. This is achieved heuristically by removing sentences that contain less than one or two of the stop-list words.

ISSUES TO CONSIDER IN PREPROCESSING

Length of Text Segments

The choice of the length of the *text segment* used for each classification n-tuple is critical for relational analysis. Part of this decision also involves whether to let these segments cross paragraph boundaries. The longer the text segment, the more concepts tend to be measured to co-occur together. This increases the

connectedness of the conceptual graph. If concepts are connected to too many other concepts, it is hard to discriminate significant and strong patterns. For extracting cognitive maps, this makes the patterns tightly grouped but unstable on repetition. Essentially there is not enough contrast in the measured relationships to get strong patterns. This behaviour is most usually observed with lexical concepts. Names tend to be much more selective with their associates. In a similar way, allowing the segments to cross paragraph or speech boundaries can add noise to the concept graph. This also makes pattern formation harder.

Dealing with Unusual Text Formats

Choice of the parameters for indexing and classification is not difficult for normal prose where the average paragraph length is greater than two sentences. Problems arise however with dialogue, plays, verse or other unusual text. Issues include that:

- For plays and dialogue, we may want to measure the relationships between characters, and between a character and all the concepts mentioned by the character. However, the speaker is normally marked at the beginning of the speech only, and they usually don't mention directly who they are talking to. So we resort to using longer text segments that not only include the whole of most speeches, but cross paragraph boundaries so that consecutive speakers form relationships. Unfortunately, this causes the lexical, or conceptual, concepts from the content of the text to be over-connected to other concepts, and the concept map becomes unstable.
- For verse, such as Blake, and other text that does not obey normal sentence punctuation and organization, is difficult for similar reasons. Sentences are either nonexistent or erratic, so the relational connectedness is noisy. Usually the best that can be done is to take each line as a sentence, or code whole paragraphs or stanzas.

VISUALISING THE NEWS

The map as shown in Figure 2 contains the names of the main *concepts* that occur within the text. Green concept labels represent proper names such as people or locations, whereas white concept labels refer to other objects, locations, actions and so on. Concepts are contextually clustered on the map. That is, concepts that appear together frequently in the text, or in similar situations, will settle close together on the map. The coloured circles on the map are *theme circles*. They aid interpretation by grouping the clusters of concepts.

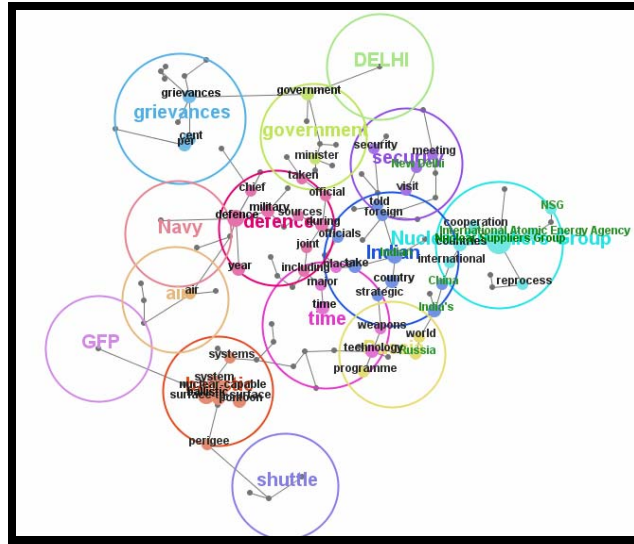


Figure 2: A concept map generated from the news events reported by major Indian Newspapers

Figure 2 displays a set of concepts that are related to a selected concept i.e. DRDO in this particular case.

Selected Concept: DRDO

Related Name-Like	Count	Relevance
Nuclear Suppliers Group	84	09%
Indian	44	05%
Rs	39	04%
India's	38	04%
New Delhi	36	04%
Navy	18	02%
Hawk	15	02%
DELHI	13	01%
GFP	10	01%
Russia	6	01%
Russian	6	01%
Sriharikota	5	01%
China	4	00%
Bidar	3	00%
North Pole	1	00%
Comptroller And Auditor General	1	00%
Army	1	00%

Related Word-Like	Count	Relevance
ballistic	433	45%
nuclear-capable	268	28%
surface-to-surface	263	27%
pontoon	221	23%
defence	155	16%
system	137	14%
developed	126	13%
systems	108	11%
perigee	108	11%
development	96	10%
technology	86	09%
project	85	09%
year	84	09%
time	61	06%
country	61	06%

Figure 3 : Display of concepts related to a particular concept

Figure 4 shows the thesaurus that is automatically built from the news for various words and concepts.

Word	Score	Word	Score
<i>Chief Controller</i>	5.15	<i>Army Commanders</i>	5.01
<i>Defence Research</i>	5.15	officer's	4.91
<i>Dr Prahlada</i>	5.07	<i>Chief Of Army Staff</i>	4.65
<i>Agni-v</i>	4.74	<i>Navy And Iaf</i>	4.57
<i>Agni-iv</i>	4.45	<i>Lt-gen Panag</i>	4.47
<i>Defence Technologies</i>	4.45	<i>Sena Medals</i>	4.47
<i>Desidoc</i>	4.45	<i>Arjun's</i>	4.36
<i>Nris</i>	4.45	<i>Dgqa</i>	4.36
<i>Awards</i>	4.37	fragging	4.36
<i>Brazilian</i>	4.37	quashed	4.36
<i>Defence Policies</i>	4.37	<i>Corps Of Signals</i>	4.24
<i>Dr. Pillai</i>	4.37	non-combat	4.24
<i>Drona</i>	4.37	sahayaks	4.24
<i>Indian Newspapers</i>	4.37	winters	4.24
<i>Information Services Division</i>	4.37	<i>Acc</i>	4.08
<i>International Conferences A...</i>	4.37	<i>Adjutant General's</i>	4.08
<i>International Relations</i>	4.37	<i>Army Chief General</i>	4.08
<i>Newspapers Clipping Service</i>	4.37	<i>Army Chiefs</i>	4.08
<i>Single Window Services</i>	4.37	<i>Battle</i>	4.08
<i>Desidoc</i>	4.28	chipping	4.08
exo-atmospheric	4.28	constable	4.08
<i>Missiles And Strategic System...</i>	4.28	cook	4.08
<i>Newspaper Clippings/datab...</i>	4.28		
<i>Research And Develop</i>	4.28		

Figure 4: A thesaurus output for "DRDO"

Figure 5 displays the various themes across which the defence related news had been reported in the newspapers in the year 2008.

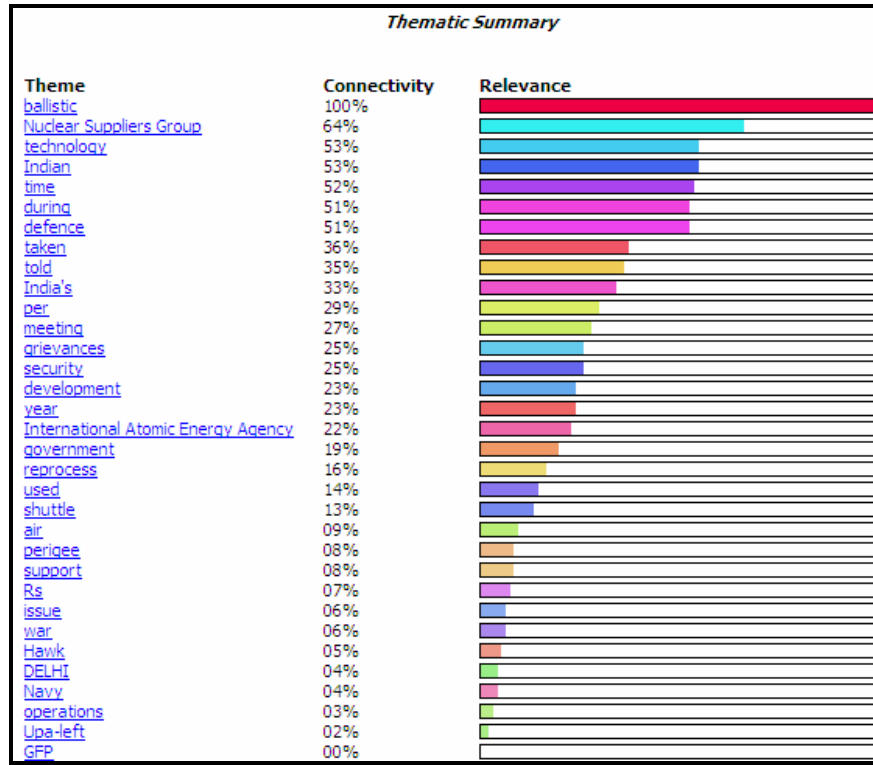


Figure 5: Themes across defence news of year 2008

CONCLUSION

Within today’s fast-changing, complex defence and security environment, it becomes increasingly important to get insights in the developments and trends that might have an impact on national security interests. Models that are used by strategic planners to support their decision-making process and policy analysis (mid and long term) and intelligence services (short term) require large amounts of data as inputs. However, collecting those data can become a technological challenge. In short, the collection of data for building and validating new models and simulation systems has to be automated or semi-automated. Key tools for this more automated approach will be improved text-mining techniques and data-visualization programmes. Such tools are becoming increasingly important parts of the modern simulationist’s tool kit. Such tools pave the way for more automated collection of the large time-sensitive data sets that are now needed and will be even more needed for simulation.

This study has the purpose to scan the news reported across various newspapers, examine the concepts and relations these have and extract the themes perceived as most useful to Indian defence and research, i.e. for foresight analysis. The

results should give a better insight in various ways to connect existing – and yet to develop – news concepts and data visualization options to analyze large amounts of collected newspaper clippings, in particular for foresight analysis. This will in effect support and improve the decision-making and policy analysis process on the strategic, operational and tactical level.

REFERENCES

1. Kroha, P., Baeza-Yates, R., and Krellner, B. 2006. Text Mining of Business News for Forecasting. In *Proceedings of the 17th international Conference on Database and Expert Systems Applications* (September 04 - 08, 2006). DEXA. IEEE Computer Society, Washington, DC, 171-175. DOI= <http://dx.doi.org/10.1109/DEXA.2006.135>
2. V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Language models for financial news recommendation", In *Proceedings of the Ninth International Conference on Information and Knowledge Management*, 2000, pp. 389–396.
3. U. Y. Nahm and R. Mooney, "Text Mining with Information Extraction", In: *Spring Symposium on Mining Answers from Texts and Knowledge Bases AAAI 2002, Stanford*, 2002
4. Murata, T. 2006. Towards the Detection of Breaking News from Online Web Search Keywords. In *Proceedings of the 2006 IEEE/WIC/ACM international Conference on Web intelligence and intelligent Agent Technology* (December 18 - 22, 2006). WI-IATW. IEEE Computer Society, Washington, DC, 401-404. DOI= <http://dx.doi.org/10.1109/WI-IATW.2006.134>
5. Allan, J., "Topic Detection and Tracking", Kluwer Academic Publishers, 2002
6. Weber, R.P. (1990) *Basic Content Analysis*. Newbury Park, Calif.: Sage Publications, 2nd ed.
7. Dubinko, M., Kumar, R., Magnani, J., Novak J., Raghavan P., Tomkins, A., "Visualizing Tags over Time", *Proceedings of WWW2006*, 2006.
8. Andrew E. Smith, *Leximancer Manual 2008*, <http://www.leximancer.com/>.
9. Kurtz, A. J. and Mostafa, J. 2003. Topic detection and interest tracking in a dynamic online news source. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries* (Houston, Texas, May 27 - 31, 2003). International Conference on Digital Libraries. IEEE Computer Society, Washington, DC, 122-124.
10. Watters, C., & Wang, H. (2000). Rating News Documents for Similarity. *JASIS*, 51(9), 793-804.
11. Krestel, R. & Mehta, B. (2008), Predicting News Story Importance Using Language Features., in 'Web Intelligence' , IEEE, , pp. 683-689 .
12. G. M. D. Corso, A. Gull'ı, and F. Romani. Ranking a stream of news. In *WWW'05: Proceedings of the 14th international conference on World Wide Web*, pages 97–106, New York NY, USA, 2005. ACM.

13. Y. Hu, M. Li, Z. Li, and W.-Y. Ma. Discovering authoritative news sources and top news stories. In H. T. Ng, M.-K. Leong, M.-Y. Kan, and D. Ji, editors, *AIRS*, volume 4182 of *Lecture Notes in Computer Science*, pages 230–243 Springer, 2006

TRAINING PROGRAMMES FOR NEWSPAPER LIBRARY PROFESSIONALS IN INDIA

Dr. G. Sivaprasad

ABSTRACT

Library and Information Science as a discipline and profession concentrates on fruitfully exploiting published information of the society. Education and training of library professionals can help an organization improve the functions and activities of newspaper libraries. The paper describes different training programmes for newspaper library professionals in India and also mentions various training programmes conducted by NISCAIR, SDC, TISS, NIRD, DELNET, SALIS, NCSI (IISc), SALIS, and IFLA-OCLC etc.

Keywords: Training Programmes, Newspaper libraries, Library professionals

Knowledge is communicable to the extent that the recipient is willing to listen to and assimilate for the day-to-day activity. After the selection and placement, each employee should be given training according to the nature of work assigned for doing a particular job. It improves the attitude and modifies the behavior of employees.

According to Dale S Beach, training refers to “the organized procedure by which people learn knowledge and improve their skill for a definite purpose”. Training programmes are desirable for an organization and its employees. Training is useful to acquire knowledge, skills, and capabilities of an individual employee to perform specific jobs are enhanced.

Education and training of newspaper library professionals can help an organisation improve the functions and activities of its library, which will enable the organisation to perform better. Education is learning to know, and training is learning to do or practice. Education provides the necessary facilities to think and evolve theories and methods of planning, operations and other related aspects with respect to the growth and development of an organisation. Training is the practice of applying methods and techniques to execute plans and programmes effectively. Every training programme should be aimed at bringing about positive change in following:

- Knowledge and skill addition
- Change in attitude and behaviour
- Improvement in the performance of work
- Improvement in organizational climate
- Affects the personal growth to minimise the resistance to change

Training programme should be organized in a systematic way to provide maximum benefit to both the library professionals and also to the Newspaper libraries. Training may be of several kinds:

- Orientation course

- Refresher course
- In-service training
- Workshops in specific areas
- Short-term courses in specific areas

The impact of Information Technology on library professionals has become deeply rooted and demands continuous improvement of professional skills of the library professionals,

TRAINING PROGRAMMES IN INDIA

In India, the frontier institutions in coordinating training and development programmes for library professionals chiefly include the National Institute of Science Communication and Information Resource (NISCAIR), SAARC Documentation Centre, Tata Institute of Social Sciences (TISS), National Centre for Science Information (NCSI), National Information System for Science and Technology (NISSAT), Documentation Research and Training Centre (DRTC), DELNET, along with other professional organisations like Indian Library Association (ILA), Indian Association of Special Libraries and Information Centres (IASLIC). The University Grants Commission (UGC) also regularly organises refresher courses through Academic Staff Colleges for the college and university library professionals and faculty members for refreshment of professional and practical knowledge.

NISCAIR, NEW DELHI

NISCAIR is a premier institution in information science, technology and systems in India. It is set up under Council of Scientific and Industrial Research (CSIR), for providing information and documentation services both at the national and international level. NISCAIR conducts following training programmes for library and information professionals:

- Information Technology for Information Management (for freshers)
- Library Automation and Resource sharing
- INTERNET access and online Information Retrieval
- Scientific paper writing
- WINISIS
- Electronic Publishing
- Designing and printing

SAARC DOCUMENTATION CENTRE, NEW DELHI

SAARC Documentation Centre is a regional centre of the SAARC. Since 1994, SDC has been conducting a number of training programmes which include short-term courses on Information Technology for Information Management, attachment training course, seminars, workshops, etc. Participants are drawn

from all SAARC nations. The training content and schedule for the attachment training programmes are specifically designed for each individual participant or group of trainees taking into account the professional background and needs.

NISSAT, NEW DELHI

NISSAT has also supported a variety of short-term courses in the areas such as:

- Application of Computers in Library and Information centres
- Use of Personal Computers and CDS/ISIS
- WINISIS and Advanced courses on CDS/ISIS
- Common Communication Format
- Internet
- HTML
- Web Design etc.

NISSAT has been organising these courses through different agencies. One-week programmes are generally organised through professional bodies and others academic institutions like NISCAIR, DRTC, universities, etc.

INFLIBNET, AHMEDABAD

The University Grants Commission also organises various training courses through INFLIBNET (Information and Library Network) at Gujarat University Campus, Ahmedabad. The following training programmes for library professionals have been initiated by INFLIBNET:

- Computer applications to library and Information services (CALIS)
- Workshop on Automation and Networking of university libraries under INFLIBNET Programme (WANULIP)
- On site training
- ILMS training
- Regional training programmes on library automation (IRTPLA) for librarians of colleges, universities and research libraries.

To provide a platform for such an interaction, INFLIBNET initiated CALIBAR in 1994, to bring together university library professionals interested in library automation and networking.

DRTC, BANGALORE

Documentation Research and Training Centre (DRTC) organises DRTC-ISI workshop on knowledge transactions, data mining techniques and decision support systems, library automation, Internet for librarians etc.

TISS, MUMBAI

Tata Institute of Social Sciences (TISS) conducts 'workshops to train library and information professionals in Information Literacy skills', 'social science research data analysis writing' and also associated with SAARC Documentation Centre conducting workshop on 'Design and development of digital libraries using open source software'.

NCSI, BANGALORE

National Centre Science Information (NCSI) conducts training courses for library and information professionals. National Institute of Rural Development (NIRD) organises training course on 'Web design and Internet information resources'.

NIHFW, NEW DELHI

National Institute of Health and Family Welfare also organises a training course on Information Technology application for information management in medical libraries.

UGC, NEW DELHI

University Grants Commission (UGC) organises various refresher courses for library professionals and faculty members through Academic Staff Colleges in various universities.

SALIS, CHENNAI

SALIS has framed its objectives by considering and accommodating the aspirations of its members particularly young and upcoming library professionals.

SALIS offers the following training programmes:

- Library Automation
- Database Creation and Maintenance
- Application of ICT in Libraries
- Knowledge Management
- Web Designing
- Institutional Repositories
- Information Literacy Programmes

Other Societies like Indian Library Association (ILA), Indian Association of Special Libraries and Information Centres, Indian Council of Social Science Research (ICSSR) organise NISSAT supported courses in different parts of the country. ILA and IASLIC also arrange a number of training programmes for library professionals.

INTERNATIONAL TRAINING PROGRAMMES

IFLA

IFLA New Training Programme for Library Association Development

IFLA has announced the Building Strong Library Associations (BSLA) programme, to be delivered under the Action for Development through Libraries programme (ALP). The programme is currently in development stage and will be launched in September 2010. The BSLA programme has been developed with extensive inputs from some of the library world's most experienced library association experts, including current and past library association presidents and executives from all over the world. The comprehensive programme offers a strategic and coordinated approach to capacity building and sustainability of library associations for the benefit of associations, libraries, and their communities. Training contents in the programme include support to develop the effectiveness of library associations in strategic planning and financial management, organisational sustainability, developing strategic relationships and fundraising, and advocating on behalf of the profession and library users.

The Building Strong Library Associations programme will consist of:

- The Building Strong Library Associations training package - a modular, customisable training package
- IFLA's existing policy training packages (for example, workshops on the IFLA Internet Manifesto or Access to Public Health Information through Libraries)
- Mentoring and activities (for example, study tours)
- Advice and help on working with other associations (for example, to form regional consortia)
- An online platform to provide remote access to materials and knowledge
- The programme can be customised and translated to meet the needs of different library associations.

ALP small projects, Building Strong Library Associations projects

The IFLA Action for Development through Libraries Programme (ALP) is announced for project applications. ALP supports projects in training and development in developing countries and emerging economies with an emphasis on capacity building and train-the-trainer or cascading delivery. Project applications can be made to the IFLA ALP Programme for support to library associations, groups of libraries, or others who want to develop co-operation, knowledge and services within the special programme areas of ALP.

The types of projects that ALP supports include:

- One-off workshops and training activities, delivering existing IFLA's policy-based training materials (for example, training materials developed by FAIFE)
- Workshops, training and capacity building projects based on ALP focal areas

- Longer-term capacity building activities and training designed to support the development of the sector in a country, for example through the Building Strong Library Associations Programme

The Jay Jordan IFLA/OCLC Early Career Development Fellowship Programme:

This programme, jointly sponsored by the American Theological Library Association, the International Federation of Library Associations and Institutions (IFLA) and OCLC, provides early career development and continuing education for library and information science professionals from countries with developing economies. On an annual basis, up to six library and information professionals, are selected for participation in this intensive five-week Fellowship programme. Four weeks are based at OCLC's headquarters in Dublin, Ohio, USA; one week is based at OCLC in Leiden, Netherlands. The programme gives Fellows opportunities to meet with leading information practitioners, visit libraries, and explore topics including information technologies, library operations and management, and global cooperative librarianship.

IFLA Action for Development through Libraries Programme Core Activity (ALP):

TRAINING ATTACHMENTS ASIA AND OCEANIA

The purpose of the programme is to further the library profession, library institutions and library and information services in developing countries. In many developing countries the status of the library and information profession needs to be strengthened. This can be done among other things by continuing education programme, especially in areas striving to keep up with the rapid pace of development in library and information services. Consequently, this is a priority area to which ALP addresses itself. The four-week training attachments are designed to afford an opportunity to middle level library and information workers from developing countries who wish to gain experience and hands-on exposure from more advanced environments. The areas identified for such training are library management, information technology, preservation, rural librarianship and literacy work.

An International Training Programme on "INFORMATION": STIMULATE

The programme has been approved by the Flemish Interuniversity Council (VLIR) and is sponsored by the Belgian Government and it is aimed primarily at persons with a university degree (Bachelor or Master), who work in universities, information and documentation centres, and libraries, including of course university libraries, and who have a few years of practical experience. It offers a stimulating learning environment to information experts in the area of science and technology, who are in the early phase of their career. The aims are to sharpen their skills in collecting, storing, retrieving, presenting and managing information. This International training programme can be of great benefit to the

teaching and research activities going on in their institutes and to the further development of their organisation and region.

CONCLUSION

Training is activities that involve development of hidden talents of individual participants and use it for the benefit of newspaper libraries. Before designing a training programme, the training institute should identify the need for the same, then design the programme keeping in view various factors like, training period, methods. After organizing training programmes, each programme should be evaluated in different areas - deficiencies have to be identified and rectified. It can be concluded that the training programmes for Newspaper Library and Information professionals can of significant value only when they contribute to the library excellence and are carried out after a diagnosis of individual and users needs. The deep-rooted impact of Information Technology on library professionals has kept them confronted with great challenges, which can be faced successfully only if the library professionals become careful about their skills and abilities and try to improve themselves through continuous involvement in various programmes of education and training.

REFERENCES

1. Anand Pawar (L) and Chandra Mouli (S). Impact of training on University and college teachers: an empirical study. *University News*, Vol. 46, No. 49, 2008, p14-20.
2. Rajasekhar (HM) and Chandar (KM). Usefulness of refresher courses in Professional development. *University News*, Vol.40, No.5, 2002, p1-3.
3. Das (AK) and Dutta (B). Training and Development programmes for library and information professionals. *University News*, Vol.40, No.1, 2002, p11-13
4. <http://www.niscair.res.in/>
5. <http://www.sdc.gov.in/News/sdcnews.jsp>
6. <http://www.inflibnet.ac.in/>
7. <http://drtc.isibang.ac.in/DRTC/>
8. <http://www.tiss.edu/>
9. http://www.ncsi.iisc.ernet.in/workshops_semi.php
10. <http://www.ugc.ac.in/>
11. http://www.nihfw.nic.in/html/Training_Activities.htm
12. www.nird.res.in
13. http://autolib-india.net/salis/salis_programmes.asp
14. www.ifla.org

ARCHIVING NEWS IN A UNIVERSITY USING OPEN SOURCE SOFTWARE

Surendran Cherukodan

and

Dr. Sheeja N.K.

ABSTRACT

Newspapers cover a large amount of information everyday on topics of varied interests. To a university, newspapers are essential components of communication as they cover various happenings in a university. These items of information are neither stored properly nor put in retrieval systems for future use. The news and views appeared in newspapers can effectively be organized in a digital library making use of open source software. The CUSAT digital library (<http://dspace.cusat.ac.in/dspace/>) has organized some news items that appeared in local newspapers about the university under a special community named "CUSAT-News". This article describes the methods of collecting, selecting, organizing, providing access and preserving news items required by a university using DSpace open source software.

Keywords: News archiving, Digital libraries, Open source, free software, DSpace

INTRODUCTION

Newspapers occupy an important place among various information sources in a library. They satisfy the requirement for information on recent events occurred at local, regional, national and international levels. They popularize understanding and awareness on numerous issues. They are at the reach of majority of population in any society. They take the multiple roles of informing, investigating, interrogating, trend setting, leading and recording social life. Hence, providing access to news and views is fundamental to libraries and information centres.

To a university community, newspapers are essential source of information on topics ranging from general knowledge to research specific news. They report the activities occurred in the campus. They give announcement of events, achievements and issues important to the university community one way or the other. They also publish views of experts on issues relevant to the academic community. They publish articles on several topics containing valuable information. Using newspapers is an important part of information gathering process of students, faculty and staff of any university. But apart from subscribing to a number of newspapers and providing access to the current and old newspapers, many university libraries are not doing the work of indexing newspapers, keeping them as clippings, preserving them on microfilming, or

digitizing them for keeping online. As a result, the items of information available in various print media are lost for future use. The management of news items appeared in print media is an important task for university libraries. This paper attempts to share knowledge on the strategies and techniques of managing news items in Cochin University of Science and Technology by using DSpace open source software.

REVIEW OF LITERATURE

Ojo-Igbinoba¹ conducted a study in 1988 that covered 19 out of 28 university libraries in Nigeria to find out the method of managing newspapers in these libraries. The study revealed that newspapers, for the most part, are not classified and therefore have not been fully integrated into the service provisions of university libraries. It was also noticed that newspapers in university libraries have not been exploited to their full potential. The author suggested that in order that newspapers may continue to fulfill their worthwhile roles in making knowledge more popular and more accessible; in fulfilling research needs and in contributing to the community intelligence, scientific, legal and historical study, the library's newspaper collection should be organized for maximum effectiveness because newspapers are documents of history. For a start each library should have a newspaper committee which would formulate a suitable policy on the place of newspaper collections in libraries.

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY: A CASE STUDY

Cochin University of Science and Technology (CUSAT)², Kerala is one among the top universities in India. CUSAT is academically organized into 9 faculties; Engineering, Environmental Studies, Humanities, Law, Marine Sciences, Medical Sciences and Technology, Science, Social Science and Technology. CUSAT has 29 departments of study and research offering Graduate and Post Graduate programmes across a wide spectrum of disciplines. As per the study by Gangan Prathap and Gupta³ for ranking of research performance of 67 Indian engineering and technological institutes during 1999-2008 using data from SCOPUS database, CUSAT bagged the 10th Rank with 1625 research papers published during the period. Among the universities, CUSAT has come up third on the list. Among the wide variety of innovative academic practices, Free/ Open Source Software (F/OSS) implementation is a priority area in CUSAT. CUSAT is running a digital library service using DSpace⁴ open source software. Giving access to news items relevant to CUSAT community is an important area in the digital library framework. Figure 1 shows the screenshot of CUSAT Digital Library Home page

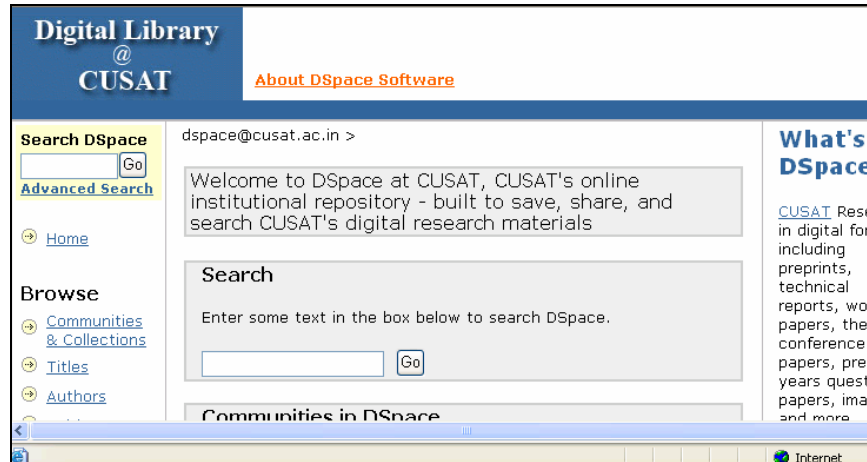


Figure 1: CUSAT Digital Library Home Page

ABOUT DSPACE

DSpace is the most widely used open source software for building institutional repositories/ digital libraries across the world by universities and research institutions for managing digital and digitized research materials. Developed by MIT Libraries and Hp Labs in 2002, DSpace has also been applied for subject based repositories, dataset repositories and media based repositories. DSpace accepts all manner of digital formats that include documents, books, theses, data sets, computer programmes, multimedia publications, administrative records, published books, overlay journals, bibliographic datasets, images, audio files, video files, reformatted digital library collections, learning objects and web pages. The DSpace software can be freely downloaded from the sourceforge⁵ open source software repository. The code is currently licensed under the BSD open source license. Any organization has the freedom to fully customize the look and feel of DSpace.

The structure of DSpace is composed of Communities, Sub Communities, Collections and Items. A community is the highest level of the DSpace content hierarchy. They correspond to an organization's departments or schools. In a University, various teaching departments form a community as shown in Figure 2. A department may have several components: faculty, library, lab, special centres, projects etc. This can be matched to the concept of Sub Communities. A sub community may have its own sub communities for example; a library may have different sections as sub communities. Collection denotes the particular group of documents in a community/sub community. Eg. The set of articles written by a teacher in a sub community of faculty constitute a collection. Items denote individual articles/ files in a collection.

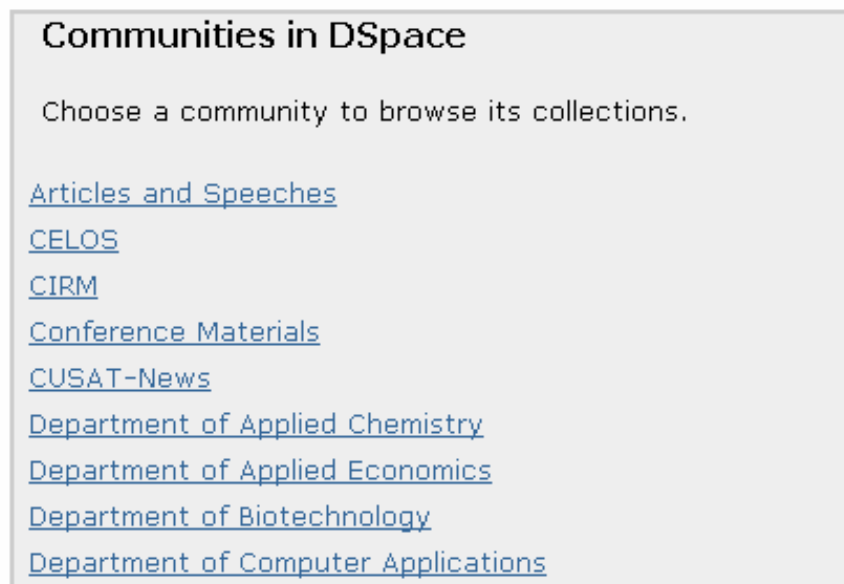


Figure 2: Communities in DSpace

NEWS IN DSPACE

CUSAT has organised in DSpace some news items appeared in print media that cover issues related to the University. It is available in a Community called “CUSAT-News”. This is made possible by selecting the item from the print media, scanning the item as a pdf file, assigning some key words, and uploading the item to the community. Figure 3 shows the news items in CUSAT Digital Library. If we click on one entry, it further displays the title, author, key words, the date of uploading, abstract given, collection name, etc. The figure 4 shows the description of one news item.

Issue Date	Title	Author (s)
2-May-2009	CUSAT developing project to protect State's coastline	Media
2-May-2009	Honour for Madhusoodana Kurup	Media
5-May-2009	Jagdish Arora at CUSAT	Media
7-May-2009	Lalu Prasad Yadav at CUSAT	Media
21-May-2009	National Conference on Copyright Law	Media
1-Apr-2009	News on CUSAT Budget 2009-2010	Media
3-Apr-2009	News on CUSAT Budget 2009-2010 Mathrubhumi Daily	Media
3-Apr-2009	News on CUSAT Budget 2009-2010 The Hindu Daily	Media
5-May-2009	Resul Pookutty at CUSAT	Media

Figure 3: News items in CUSAT Digital Library

Title: CUSAT developing project to protect State's coastline

Authors: Media

Keywords: Coastal zone management
Rubberised Coir
sea erosion
Joy Job Kulavelil

Issue Date: 2-May-2009

Abstract: News item appeared in the Hindu Daily on 30-4-2009

URI: <http://hdl.handle.net/123456789/2440>

Appears in Collections: [News](#)

Files in This Item:

File	Description	Size	Format
CUSAT developing project to protect State's coastline.pdf		50 kB	Adobe PDF

[View/Open](#)

Figure 4: Description of one news item in DSpace

For viewing the full text there is option for **view/open**. When clicking on it, the scanned news content is opened as a pdf file. The figures 5 and 6 show the display of actual news content in DSpace. The name of the newspaper and date is written on the news item before scanning so that those who save and print it need not trace this information in the abstract section. The clarity of the news item in pdf depends on the quality of the original. The performance of the scanner is also a factor for image quality. However, it is better to keep the actual physical features of the original as far as possible.

address <http://dspace.cusat.ac.in/dspace/bitstream/123456789/2440/1/CUSAT%20developing%20project%20to%20protect%20sta>

Save a Copy Print Email Search Review & Comment Sign

Select Text 100%

Cusat developing project to protect State's coastline
THE HINDU 30-4-2009

G. Krishnakumar

KOCHI: Cochin University of Science and Technology is developing an ambitious plan to save the ecologically-sensitive coastal belt by involving the traditional coir and rubber farmers in the State.

The objective of the project is to evolve a multi-disciplinary coastal zone management by incorporating hard and

- Rubberised coir will be used to check sea erosion
- Farmers to benefit from demand for coir and rubber

the ailing coir and rubber sectors in the State. Dr. Kulavelil said that the university scientific team, led by its Registrar N. Chandramohankumar, has suggested that the rubber and coir farmers will get support, as the project requires several tonnes of coir and rubber during its implementation.

It will also promote the plantation of rubber and coconut trees in the State, he

Figure 5: News in PDF file in DSpace

MANAGING NEWS IN DSPACE IN A UNIVERSITY

DSpace structure maintains a decentralized pattern for knowledge organization. The organizational structure of a university can be easily mapped in DSpace.

Here each community can build data relevant to them. The first task of archiving news in DSpace is the creation of a Sub Community named “News and Views” in each Community. Within this Sub Community, we can create “Collection” with suitable name. In each collection we can store individual news items. The responsibility of collecting/identifying required news and views from newspapers can be assigned to a staff member in each Community. The process of scanning the item, saving it in pdf, the work of uploading the item to Community collection, providing abstract, giving keywords and keeping file backups can be done either by a central agency or by the respective Community.

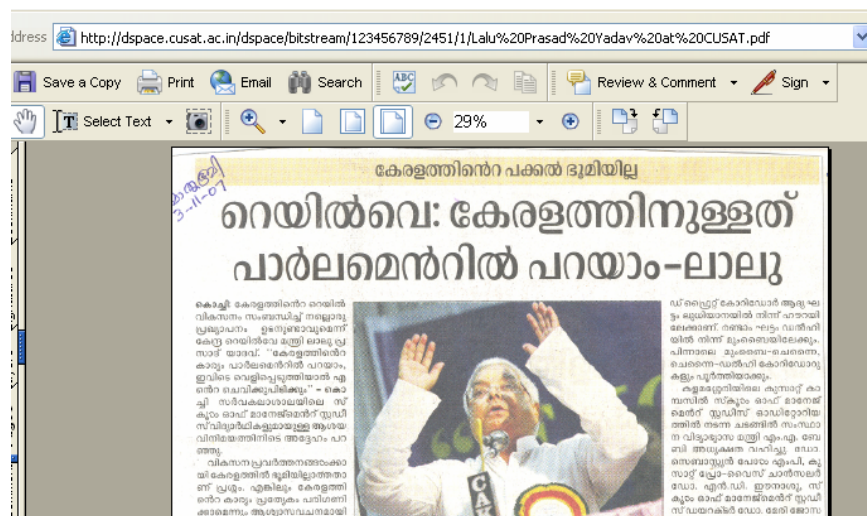


Figure 6: News Item in local language in DSpace

For systematic course of collecting and storing news in a university, there need to be a policy for the work. The information appeared in newspapers may lack authority, clarity and relevance. Some news items may be critical and controversial. Archiving of these items may not be recommended by the authority. These issues can be solved only through a sound institutional policy for news archiving.

BENEFITS OF USING DSPACE FOR NEWS ARCHIVING

News and views appeared in print media may not be read by many users as the newspaper editions are region specific. The news items being preserved in DSpace can be viewed and used online by all. It shall enhance access to news and views. Getting top search result in Google is another advantage of DSpace. The Figure 7 shows the screenshot of Google results when searching “news at CUSAT”.

[DSpace at Cochin University: News And Views](#)

News And Views. Community home page. In: All of DSpace, **News And Views**, Important Events.

Search for, or browse. Collections in this community...

dspace.cusat.ac.in/dspace/handle/123456789/2073 - [Cached](#)

Figure 7: Google result for searching “news at CUSAT”

The retrieval of pertinent items from large number of files can be managed by assigning relevant keywords to each file. The archiving of news shall help the process of teaching, learning and research in a university. If the archive can be accessed online, the educational and informative contents can be used by the wide public. Since DSpace is having a large community of developers and users the world over, the institution shall get support in matters of using and developing DSpace platform for news archiving.

CHALLENGES OF NEWS ARCHIVING

The successful archiving of news in a university by using DSpace open source software requires the continuous effort from the faculty and library professionals in monitoring and capturing news and views by their respective community. They must also be alert in indexing, abstracting and converting the documents in the bit stream accepted by Dspace software. Technical challenges include the support of the university in procuring required hardware, customisation, server maintenance, effective mechanism for back-up and coordination of document management in the system.

CONCLUSION

The process of archiving news and views is an essential activity in a university. It shall improve the way people use news and views. It shall help the university to store the required news items in an organized, secure, and searchable archive and preserve it for long time. The use of open source software platform for news archiving helps to save software cost. The support of experts in the field is also available through email forum. The practice of using DSpace open source software for news archiving is a simple, cost effective and reliable method for all types of institutions.

ACKNOWLEDGEMENT

The authors acknowledge the help and support of G.Santhosh Kumar, Lecturer, Department of Computer Science, CUSAT and CUSAT Digital Library administrator who is currently the south Indian Ambassador of DSpace.

REFERENCE

1. Ojo-Igbinoba, M.E The Role and Management of newspapers in Nigerian libraries, *International Library Review* (1991) 23, 83-90
2. <http://www.cusat.ac.in>
3. Prathap, Gangan, and Gupta, B.M “Ranking of Indian engineering and technological institutes for their research performance during 1999-2008”, *Current Science* Vol.97 No.3 10 Aug 2009. pp 304-306.
4. <http://www.dspace.org/>
5. <http://sourceforge.net>

PRESS CLIPPINGS INFORMATION SERVICE OF THE JAWAHARLAL NEHRU UNIVERSITY LIBRARY: A CASE STUDY

Mahesh Chand

INTRODUCTION

The Jawaharlal Nehru University (JNU) was established in the year 1969 for nurturing the higher education in India and to take it to further heights. The JNU campus is located on 1000- acre area of the Aravali Hill range of South Delhi. Part of the campus hosts dense forests, sustaining a birdwatcher's paradise and some forms of wild life.

THE JNU LIBRARY

The Jawaharlal Nehru University Library is a Knowledge Centre which has rich resources mainly in Social Sciences, Humanities and Sciences. It is a nine-storey tower building and has a carpet area of about one lac sq. ft. It is situated in the midst of the academic complex of the University and is the hub of all the academic activities of the University. The JNU Library incorporates the library collection of the prestigious Indian School of International Studies which was later merged with Jawaharlal Nehru University. The JNU Library is a depository of Government publications and publications of international organisations like WTO, WHO, European Union, United Nations and its subordinate agencies etc.

For meeting the research requirements of the Faculty and Research Scholars of the University, the JNU Library started several premium information services; one of these services is Press Clippings Information Service. Initially this service was meant for the Faculty and Scholars of the School of International Studies (SIS). The SIS is a premier institution in India for the study of international relations and area studies. The SIS is promoting the study of international relations as an academic discipline in India and in advancing knowledge and understanding of international affairs in an inter-disciplinary perspective. To cater to the diverse and current information requirements of the Faculty and Research scholars of SIS and also to supplement the monograph and serial collection, the Press Clippings Information Service was started in August 1974.

IMPORTANCE OF NEWSPAPER INFORMATION

The main conveyor of news is the daily newspapers. Though there are many other sources of news like magazines and books in print media, radio and television channels and Internet in electronic media, words of mouth and telephone and wireless communications etc, no other source ever could come near the newspaper in importance. For years, newspapers have provided and preserved detailed records of topics, people, institutions, issues and events. Newspapers are filled with valuable records, facts, statistics, analysis, references, quotations and much more. Today, a wealth of credible local, regional, national and international newspaper coverage continues to provide people worldwide with access to the essential

information they require every day. Consequently, newspaper archives remain among the most widely used resources in all types of libraries.

PRESS CLIPPINGS INFORMATION SERVICE

The JNU Library provides a number of user services, and the Press Clippings Information Service is one of the premium services being offered. The press clippings information service is provided by many libraries and information centers in India but at the university level the press clippings information service of the JNU Library is the most comprehensive service in India.

The Press Clipping Information Service was started in 1974 to cater to the information requirements of the faculty and scholars of School of International Studies on topical issues. It collects all important, relevant and up-to-date news items, editorial comments and articles on developments in the international and bilateral relations, political, economic, socio-cultural, natural hazards etc. At present, the clippings are taken from 18 English newspapers published in various parts of the country and two newspapers from overseas. The Press clippings are maintained chronologically in subject folders and stacked in proper sequence according to the Colon Classification scheme.

The following services are being provided to research scholars:

Reading Room

The Press Clippings Section of the JNU Library is providing reading room for users. The section is open for scholars from Monday to Friday, 09.00 AM to 05.30 PM, except on national holidays. Access is available for research and reference for the University faculty and scholars; the guest scholars can use the services with the introductory letter from their respective departments/ institutes and permission from the University Librarian.

Photocopy Services

Photocopy services are available only from the loose press clippings kept in the pigeon holes. The press clippings from folders can't be copied for the safety, security reasons and the short life of the paper.

Information and advice

Professional staffs are available to give advice to the scholars on:

- how to use the press clippings collection
- to assist users with their enquiries
- provide reference, SDI, CAS services

SCOPE

The Press Clipping Information Service was started on a modest scale with the objective of speedy disposal of references on current topics received from Faculty and Research Scholars from School of International Studies. The core subject coverage initially was international and bilateral relations, foreign policy, economic conditions, international organisations, women studies, natural disasters etc. The scope of the press clippings information

service has considerably increased over the years and accordingly the coverage has expanded. The Service collects all important, relevant and up-to-date articles, editorial comments and articles on developments in the legislative, political, economic, socio-cultural fields. The subject range covers almost all the social sciences disciplines like – economy, industry, labour, agriculture, religion, politics, elections, land reforms, centre state relations, trade, tourism, national security, environment etc.

METHODOLOGY

The news paper articles are selected from the newspapers on the basis of subject coverage above and which have a research value for the University. The selected press clippings are broadly classified according to Colon Classification system; after classification each press clipping is pasted on a base paper and stamped with name of newspaper, its edition and date of the publication. (Figure 1). After pasting and stamping, the press clipping are sorted according to subject and kept in pigeon holes. From pigeon holes, the scholars can consult and Xerox press clippings. After four weeks the loose press clips are filed chronologically in the subject folders and press clippings can't be photocopied. Due to safety and life considerations, press clippings folder can be consulted in the reading area only attached to the Section. At present, the Press Clippings Section is housed at the seventh floor of the JNU Library Building.



Figure 1: Sample News Paper Clippings

THE COLLECTION

The press clippings collection consists of over 12 lakhs of press clippings on different topics of social sciences and area studies. Press clippings are

maintained chronologically in subject folders according to broad class numbers based on the Colon Classification system. These folders are not issued out. These completed press clippings folders can be consulted in the Reading Room of the Press Clippings Section only. Finally the press clippings boxes are arranged on the shelves according to subject and area wise.



Figure 2: Newspaper Collection

USAGE

The press clippings are being used extensively specially by the faculty, research scholars of School of International School of Jawaharlal Nehru University and the research scholars of other social sciences and humanities discipline. Guest research scholars and social scientists from all over the country from different universities and institutes of higher education are also visits the JNU Library for using this unique collection.

DIGITISATION OF PRESS CLIPPINGS COLLECTION

The enormous advantages of the new digital media in terms of both supply and demand have brought about a veritable revolution in modern times. Also the short life of the newspaper and damage caused to the press clippings collection by continuous usage led to planning digitization of press clippings collection. In the traditional form of this collection, it was becoming difficult to maintain and provide service. To safeguard the collection and provide the better access to scholars, the JNU Library has started digitisation process and is preparing a database of this collection.

Out of the different options available for digitisation, the JNU Library opted 'VIRTUA' software; since the JNU Library is using Virtua, the integrated library management software for library and this software can also be used

for image database also. The JNU Library started using Virtua for full text image database of press clippings.

The work of back files digitisation has been outsourced to VTLS Corp. Noida. The current job is being carried out in-house by the professional staff of the JNU Library. Presently the database is having about 35000 images and is available for access by university academic community on the WEB-OPAC of the JNU Library.

The press clippings database can be accessed by title of the newspaper, date of publication, author of the article, subject headings etc.

http://172.16.21.5:8000/cgi-bin/gw_clip/chameleon/

* (the article is based on practical experience of the author)

BIBLIOGRAPHY

1. Sen Uttam. **Newspaper as a looking glass.** *Mainstream*. Vol. 47(18): 26-30.
2. <http://www.shgresources.com/resources/newspapers/world-newspaper-library/> (access on 25 January 2010)
3. Dans, E.(2000). **Internet newspapers: are some more equal than others?** *The International Journal on Media Management*. Vol 2 (1): 4-13

List of Authors

Ajit Pyati

Assistant K. Pyati
Faculty of Information and Media Studies
University of Western Ontario
London, ON, Canada

Alaka Buragohain

(Retd Professor, Dept of Lib & Inf Sc, G U)
Honorary Director, Media Archives

Anita Pujari

General Manager – Archives & Syndication
Research N Archives
DNA – Daily News & Analysis
E-mail : anita@dnaindia.net

Ashis Biswas

Librarian
Gurunanak Institute of Dental Science and Research
(A Post Graduate Dental College under DCI)
Kolkata, W.B.
E-mail : lisbashis@gmail.com, biswas.ashis@yahoo.com

Avijit Chakrabarti

Assistant Librarian
Barasat College
E-mail : avijitchakrabarti@yahoo.co.uk

Biplab Chakrabarti

HoD, Department of Library & Information Science
University of Calcutta
E-mail : bcliscu@gmail.com, bclis@caluniv.ac.in

Cathy Pilgrim

Director
Digitisation and Photography Branch
National Library of Australia
Canberra Act 2600
E-mail : cpilgrim@nfa.gov.au

Deborah Novotny

Head of Preservation
Collection Care
British Library
96 Euston Road
London, U.K.
E-mail : deborach.novotny@bl.uk

Debora Cheney

The Larry and Ellen Foster Communications Librarian
and Head, The News and Microforms Library
The University Libraries
The Pennsylvania State University
office: W-24 Pattee Library, USA
E-mail : dcheney@psu.edu

D. R. Gupta

Sr. Reprography officer
Indira Gandhi National Centre for the Arts,
11, Mansingh Road, New Delhi
E-mail : drg_ignca@yahoo.co.in

Else Delaunay

Bibliothèque Nationale De France
DCO/Droit, Economie, Politique
Quai François Mauriac
75706 Paris Cedex 13, France
E-mail : else.delaunay@bnf.fr

G. Sivaprasad

Lecturer in Library Science
MVGR Degree College, Bhattiprolu Guntur (Dt.)
E-mail : sivaprasad_gorantla@yahoo.co.in

Gopa Ghosh

Library & Information Assistant
National Library, Kolkata
E-mail : gopa_gh@yahoo.co.in

Joshy Kuriakose

Center for Development of Digital Libraries (CDDL)
Indian Institute of Management Kozhikode (IIMK)
IIMK Campus P.O., Calicut, Kerala

Keith Rajecki

Industry Solutions Architect
Global Education & Research
Sun Microsystems, Inc., USA
E-mail : keith.rajecki@sun.com

Lavji N. Zala

MICA, Ahmedabad

Majlis Bremer-Laamanen

Director, The National Digitization Centre
Centre for Microfilming and Conservation
The National Library of Finland
E-mail : majlis.bremer-laananen@helsinki.fi

M.G. Sreekumar

UNESCO Coordinator,
Greenstone Support for South Asia
Librarian & Head
Center for Development of Digital Libraries (CDDL)
Indian Institute of Management Kozhikode, Kerala
E-mail : mgsree@iimk.ac.in

M. Natarajan

Scientist, Niscair
New Delhi
E-mail : m_natarajan@hotmail.com

Mahesh Chand

Assistant Librarian
Jawaharlal Nehru University
New Delhi
E-mail : mchand@mail.jnu.ac.in

Mohamed Musthafa. K

Research Associate
Dept. of Library and Information Science
Aligarh Muslim University,
Aligarh, India
E-mail : musthafa.lis@gmail.com

Mayuri Das Biswas

Librarian, Adamas International School,
Kolkata
E-mail : mails.mayuri.das@gmail.com

Ngian Lek Choh

Deputy Chief Executive and Director
National Library, National Library Board
Singapore

Naushad Ali P.M

Chairman
Dept. of Library and Information Science
Aligarh Muslim University
Aligarh, India
E-mail : naushadali.amu@gmail.com

Niraj R. Patel

MICA, Ahmedabad

Nivedita Bhattacharyya

University Research Fellow
Department of Library and Information Science, University of Calcutta
E-mail : nives_bhattacharyya@yahoo.com

O.N. Chaubey

Indira Gandhi National Centre for the Arts
11, Mansingh Road, New Delhi
E-mail : onchaubey@gmail.com

Patricia Liebetrau

Director of KISA
Digital Innovation South Africa, DISA

Payel Biswas

AIS Student, Nisclair
New Delhi
E-mail : dipalibiswas.pia@gmail.com

Prartibha Kaushik

Executive –Image Archive
Hindustan Times Media Ltd.
E-mail : pratibha.kaushik@hindustantimes.com

Pritam Gurey

Professional Assistant
Visva Bharati, Central Library, Santiniketan
E-mail : pritam.vb@gmail.com

P. Sreejaya

Center for Development of Digital Libraries (CDDL)
Indian Institute of Management Kozhikode (IIMK)
IIMK Campus P.O., Calicut, Kerala

Robert B. Allen

College of Information Science and Technology
Drexel University
Philadelphia, PA 19107 USA
E-mail : rba@drexel.edu

R. Venkata Kesavan, AGM

Times Archives and Knowledge Centre (TAKC)
The Times of India Group, II Floor, TOI Building.
D. N. Road, Mumbai

R.Biju

Center for Development of Digital Libraries (CDDL)
Indian Institute of Management Kozhikode (IIMK)
IIMK Campus P.O., Calicut, Kerala

R. Chandra

Scientist/Librarian (retd.)
Institute of Nuclear Medicine & Allied Sciences
DRDO, Min. of Defence, Delhi
E-mail : cpsrkrishnan@yahoo.co.in

Shailesh R. Yagnik

Librarian,
MICA, Ahmedabad

Sumeer Gul

Assistant Professor
Department of Library and Information Science
University of Kashmir, Hazratbal, Srinagar
E-mail : sumeersuheel@gmail.com

Sumit Goswami

Defence Scientific Information and Documentation Centre (DESIDOC)
DRDO, Min of Defence, Metcalfe House, Delhi
E-mail : sumit_13@yahoo.com

Surendran Cherukodan

Junior Librarian
School of Engineering
Cochin University of Science and Technology
Cochin, Kerala
E-mail : csura@cusat.ac.in

Shankar Marathe

Sr. Manager
Times Archives and Knowledge Centre (TAKC)
The Times of India Group, II Floor, TOI Building,
D. N. Road, Mumbai
E-mail : shankar.marathe@timesgroup.com

Salila Sreesan

Dy. Chief Librarian
Times Archives and Knowledge Centre (TAKC)
The Times of India Group, II Floor, TOI Building,
D. N. Road, Mumbai
E-mail : salila.sreesan@timesgroup.com

Sheeja N.K.

Assistant Librarian
University Library
Cusat, Cochin
E-mail : sheeja@cusat.ac.in

T. Sunitha

Center for Development of Digital Libraries (CDDL)
Indian Institute of Management Kozhikode (IIMK)
IIMK Campus P.O., Calicut, Kerala

Tariq Ahmad Shah

Library Assistant

Islamic University of Science and Technology

Awantipora. Pulwama, Jammu and Kashmir

E-mail : tariqahmadshah@gmail.com

V. Senthil

Defence Scientific Information and Documentation Centre (DESIDOC)

DRDO, Min of Defence, Metcalfe House, Delhi

E-mail : senthildrdo@gmail.com

Vijya P. Rajhansa

Head, D.L.I.S

University Librarian

University of Mumbai

Vidyanagari, Santacruz (East), Mumbai

E-mail : vprajhansa@yahoo.com

Editors' Profile



Dr. Ramesh C.Gaur

Ramesh C.Gaur is Librarian & Head, Kala Nidhi Division (Professor Grade) at Indira Gandhi National Centre for the Arts (IGNCA), New Delhi. He is the Project Coordinator of an International Bibliographic Project known as ABIA. A Fulbright Scholar, Dr. Gaur is a Science graduate with Ph.D in Library and Information Science.

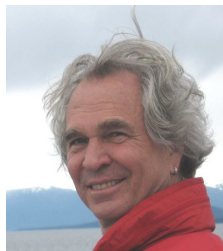
After rich exposure to corporate environment, he joined IMT Ghaziabad where he was instrumental in automating the Library and transforming it into a state-of-the-art establishment, hailed by all concerned. A feat that he went on to replicate at the famed Tata Institute of Fundamental Research, Mumbai and also at Management Development Institute (MDI), Gurgaon. Digitization and Digital Archiving of Indian Cultural Heritage is the main focus of his work at Indira Gandhi National Centre for Arts.

Dr. Gaur has visited USA, Germany, Thailand, Indonesia, Cambodia, Bangladesh, Malaysia and the Netherlands on various assignments related to Digital Library and allied areas.

A recipient of several national awards, including IASLIC- S M Ganguly Best Librarian award for year 2007, Dr Gaur was honoured with Satkal Young Librarian Award for year 2009.

He is Member of Delhi Public Library Board. He has served on various committees, task forces constituted by the Ministry of Culture, Ministry of Information & Communication Technology, Government of India including the UGC National Committee on e-theses and National Knowledge Commission Committee on Private Collections etc.

In his 20 years professional career, he has published over 30 papers, authored/edited three important books in his chosen field, and sponsored/supported organizations and causes related to Library Science.



Frederick Zarndt

Frederick has worked with historic and contemporary newspaper digitisation since computer speeds, software technology, and storage capacities and costs first made it practical. He worked with the Library of Congress on the pilot implementation of NDNP and with the University of Utah since the beginning of its newspaper digitisation programme. Frederick has experience in every aspect of digitisation projects including project requirements

development, project management, conversion operations (both in-house and outsourced), acceptance testing, and software development for production and delivery of digital data. Frederick is President of Planman's North American company and was CTO at iArchives prior to this. He has 25+ years experience in software development and is member of ACM and IEEE and a Certified Software Development Professional (CSDP). Frederick has a Master's Degree in Computer Science and Physics.



Dr. D. R. Gupta

Dr. D.R. Gupta holds a masters degree in History and a Ph.D in Social Sciences. He earned several certificates and diplomas in the field of Reprography, Care and Conservation of Archives, Record Management, Computer Science, Art & Culture and Language & Literature etc.

Starting his career as Assistant Microphotographer in National Archives of India, Jaipur in 1987, Dr. Gupta is presently with Indira Gandhi National Centre for the Arts (IGNCA), New Delhi working as Senior Reprography Officer. During his 23-year long professional career span, he has been extensively involved with several academic and research initiatives. He is an active participant in national and international seminars and workshops and has contributed more than 12 papers on contemporary subjects of national significance. His area of interest is IT infrastructure and he has got much exposure in computer hardware and software technologies especially digitization operations. Dr. Gupta is a Life Member of ILA and IASLIC.



Mrs. Kavita Gaur

Mrs. Kavita Gaur is presently working as Assistant Librarian, International Management Institute (IMI), New Delhi. A post graduate in Geography, Mrs. Kavita has also done her masters in Library & Information Science. Her major areas of interest include library classification, cataloguing and reference services. During the past 16 years of her professional career, she has contributed more than ten papers to various professional journals and conferences. She has also attended a number of National and International conferences and has acted as Rapporteur in various technical sessions in some of these conferences.