

## Digitization of Archival Collections by Libraries in Pakistan: Issues, Strategies, Challenges and Opportunities

Bushra A. Jaswal

Forman Christian College, Lahore, Pakistan.

Email: bushrajaswal@fccollege.edu.pk



In the backdrop of a case of a major archive digitization initiative completed by a Library in Pakistan to preserve a decaying last copy of an Urdu News Magazine from 18th century, the author highlights the threats to documentary

heritage in general and in Pakistan in particular. The paper gives a comprehensive discussion of the digitization process, its issues, challenges and constraints by highlighting how in Pakistan, libraries and documentary heritage are facing neglect and a threat for extinction due the environmental, economic and political conditions and human threats like terrorism. Discussing the practical and operational problems of digitization work in an environment where the archival digitization is not a common practice, the paper suggests some workable strategies and solutions to overcome the barriers and bottlenecks. The paper is a good example to educate the future digitizers in Pakistan and offers workable suggestions and many thought provoking ideas for local librarians for saving the decaying archives and rare materials held in their libraries, by collective and collaborative digitization efforts.

**Keywords** Digitization; Digital Libraries; Digital Archives; Nur-i-Afshan Newspaper; British Library; Project EAP660-British Library Endangered Archives Program; Ewing Memorial Library; Forman Christian College; Developing Countries

### Introduction

The world of information, libraries, archives and museums is going through a fast transformation!

More and more digital information is produced and delivered and electronically accessed over the Internet, by the scholarly community of the world. Despite the proliferation of digitally produced information, there still is a huge amount of historic information in print as well as other analog formats which is stored in the libraries and archives. In some cases, the older information is more useful due to the supreme value of historical evidence. The enormous volume of scholarly information being generated in digital formats could not diminish the demand of archival information in the analog formats but the limitations of usage due to their analog characteristics has created a strong need and a pressure on the libraries for digitization of their archival collections.

The newer digital formats have created a new hope for the archives and cultural heritage records to be preserved through digital surrogates for the future use as well as to save the valuable historical information from being perished with the decaying paper and obsolete technologies.

This paper reports an archival digitization project completed by a University Library in Pakistan and describing the threats to archives collections highlighting the challenges, issues, problems and bottlenecks experienced in the due course of completion of the archival digitalization project in a developing country setting.

## Digital Archives

Digital archives are the analog archival materials converted into machine-readable format using the digitization equipment, standards and techniques, with the purpose of preservation or to make them more accessible to users. The term may also be used for both the born-digital information and the information preserved in the same format for its archival value. (Verzosa, 2004)

*The archives challenge.* The documentary heritage from the past centuries, wherever it exists in the world, is facing much more challenge of extinction than it did in the past. Luckily, the awareness exists at global level that the archival materials around the globe are in danger, as stated by the UNESCO Memory of the World Programme, "Documentary heritage reflects the diversity of languages, peoples and cultures. It is the mirror of the world and its memory. But this memory is fragile. Every day, irreplaceable parts of this memory, disappear forever" (UNESCO, 2015)

### The Status of Archives Digitization in Pakistan

Pakistan has the pride of having some very old and prestigious public and academic libraries throughout the country, which are known for their rich collections and a wealth of documentary heritage of rare manuscripts, historic newspapers and archival value periodicals portraying the eventful past of the people in South Asia region. (Ahsan, 2009) reported the availability of 0.15 million manuscripts in Pakistan in Arabic, Persian, Urdu, Pashto, Sindhi and Sanskrit languages. Some are available in libraries of Pakistan while others are in personal collections.

Most of this rare material is at the verge of extinction due to the aging analog media. There is an urgent need to preserve this material by converting them to the digital formats to save the loss. Some efforts being done in this area are as follows:

*The National Library of Pakistan* holds 555 manuscripts including the rare handwritten

manuscripts in Persian language relating to history of Kashmir and microfilms/microfiches of all India Census Reports for 1911, 1921, and 1931. A Digital Information Division has digitized 10000 rare books, mostly in Persian language. The digital images are saved in static image formats in folders. (Ghayyur, 2012)

*The Punjab University Library* has the largest manuscript collection in Pakistan. It has 22574 manuscripts including 13167 Urdu, Arabic, Persian and local languages and 1871 manuscripts in Hindi and Sanskrit. Sanskrit and Hindi manuscripts are written on Palm leaves. 500 manuscripts have been digitized in image form and are accessible at library premises only. (Ahsan, 2009).

*Pashto Academy, University of Peshawar* is digitizing manuscripts in Pashto language with the help of United States Agency for International Development (USAID). There are nearly 0.15 million manuscripts in Pakistan which are in Arabic, Persian, Urdu, Pashto, Sindhi and Sanskrit languages.

*Iqbal Academy, Lahore* is digitizing fiction and poetry of Urdu literature and publishing online on the Iqbal Cyber Library Website. The site contains 1003 books from 385 contributors on 101 Subjects in 18 Languages. The books may be accessed by browsing through various indexes and may be retrieved by downloading the PDF files (Iqbal Academy, 2014)

*Punjab Public Library* makes scanned images of archives and rare materials and saves them in image format on the CDs and DVDs to provide them to the visiting users on demand.

These initiatives reveal some activity towards the digitization of archives in Pakistan. However, the efforts being done are very few and woefully deficient in terms of standardization, quality, resources and archival preservation standards. None of these initiatives provides online searching, access and retrieval to the users.

*The Sindh Archives, Karachi* started digitization during 2005. Software titled, 'Manuscript Search Program' (MSP), has been developed involving

senior professors in Library Science, experts in manuscripts and IT professionals for entire collections. Brief description/ catalogue of manuscripts available at this repository may be available online shortly through website (Sindh Archives, 2013)

According to another study, some university libraries of Pakistan are also engaged in digitization activities, however, at a budding stage. The study found that theses and dissertations were the top most priority for digitization in terms of types of material and social sciences was found the most focused area for digitization. The highest ranked digitization goals of these projects were reported as; 1) to provide access via web, 2) to increase access, and 3) preservation. (Rafiq & Kanwal, 2013). However, the achievement of indicated goals of online web access and increased access are not visible as yet at any university in Pakistan.

Additionally, some projects of archives' digitization are currently underway in Pakistan with the funding support and standardization guidance from the Endangered Archives Program of the British Library UK. These are the following:

*EAP566: Endangered Urdu periodicals: preservation and access for vulnerable scholarly resources* started in 2012. This is an ongoing major project of £52,247 for 24 months. This Project is underway at the Mushfiq Khwaja Library and Research Centre, Karachi, as an archival partner.

*EAP766: Retrieval, cataloguing and photographic imaging of rare manuscripts, Balochistan-Pakistan* started in 2014 as a Pilot project £12,276 for 8 months, by the Balochistan Archives as an archival partner *EAP660: Nur-i-Afshan archives: Perspectives on the inter-religious history of Punjab from 1873-1944* started in 2013 as a Major project award of £35,600 for 12 months, by the Ewing Memorial Library, Forman Christian College as the archival partner. This Project is now complete and published online at the Endangered Archives

Program Repository (Endangered Archives, 2015).

This paper details the EAP660 Project as a typical example of digitizing the archives of an almost extinct Newspaper from the 18th Century to discuss the challenges and to highlight the issues, and problems faced during the digitization process.

### **The Institution**

Dr. Charles W Forman, a Presbyterian missionary from the USA, founded Forman Christian College (FCC), initially known as the Lahore Mission College in 1864. In 1894 the name was changed to Forman Christian College to honor the founder. FCC was nationalized by the Pakistani government in 1972 and remained under government control until March 2003 when it was returned to its former owners, the Presbyterian Church (USA). In 2004, FCC was awarded the status of a University under a Charter of the Government of Punjab, Pakistan, and it now offers baccalaureate, masters and MPhil. degree programs. 2014 was the year of 150th Anniversary of FCC. (Forman Christian College, n. d.)

The Ewing Memorial Library was initiated at a very early stage by the Founders of FCC in the old FCC Building in Neela Gumbad, Lahore. Later, a three-floor library building was built in 1943 at the new FCC campus' Ewing, the second Principal of the College. A rich collection of the selected books and journals of that era was developed and carefully maintained by renowned librarians and scholars of the FCC throughout 150 years of its existence. The author joined this Library as a Chief Librarian in 2007.

### **Challenges and Threats to Digitization of Archives in Pakistan**

While the efforts of saving the cultural and documentary heritage of the nations is quite encouraging at a global level, a responsible Librarian sitting in a 150 years old academic institution would wonder at the volumes of decaying papers around and would think the possibilities of preserving the knowledge

reservoirs collected by the past librarians for transferring them to the coming generations.

The question is difficult to answer because the collections in the very old Libraries like the one in Forman Christian College, have weathered through the following challenges:

*Fragility and obsolescence* The use of fragile paper and obsolete physical formats cause threats to the very existence of the archives. The fragile acidic paper of old times was used by the primitive printing presses of the subcontinent which is the major cause of decay from within. Then, the books and journal that are now considered as rare were not produced with this purpose. Nobody realized their importance for a long period until they were suddenly found as a last available copy in the world. By that time, it was already too late to restore the damaged material. At times, the information on older storage media has become unusable due to non-availability of the required technology.

*The environment* Pakistan has severe weathers with heavy monsoons as a regular feature. The old Library buildings are mostly older than the paper they are housing and are unable to face calamities like floods and torrential rains. There hadn't been enough arrangements in the past to save the paper collections from humidity, heat and pests. An ongoing preventive and remedial treatment is a must, but not possible without strong institutional support.

*Scanning equipment and expertise* Economic conditions in this part of the world have never been good. Hence, the publications from the last centuries were printed on a very low quality paper to save the cost. This paper deteriorated fast and in most cases has already passed its usual survival time. It had become too brittle to endure human touch. Handling this material requires special expertise and expensive equipment for scanning which makes its digitization, a difficult, cumbersome and expensive task. Securing funds and time required for this expensive and laborious task is a big challenge.

*Lack of resources and neglect* Libraries had never been a priority for any government in Pakistan. Therefore, library budgets had been too low to afford air-conditioning or any other forms of humidity and temperature control to save the damage to the paper-based collections. With the passage of time the paper documents crumbled and became unable to retain their value as knowledge sources.

*Knowledge of archives*, the non-availability of professional training to educate librarians and archivists resulted into lack of appreciation of the rare materials. Precious information materials are sometimes lost just because of library staff's inability to recognize the archival value of the items.

*Funding* Digitization is a good solution for librarians to play their part in the transfer of human knowledge by preserving the decaying analog information in digital formats. However, scanning and digitization is a slow and highly expensive process. Non availability of adequate funding for this activity is a major problem.

*Access and retrieval systems* The expertise and skills to establish and operate the Hi-speed information technology infrastructure are required for a hi-speed online digital repository system for offering online access. Moreover, the retrieval of digitized documents is not available in libraries of Pakistan. Often money stands more valuable than the information.

*Skilled manpower* required for digitization and digital repository operations are not available in Pakistan. Library schools in Pakistan offer minimal courses on Digitization, Digital Library Development and Online Repositories. Library associations, professional groups and academia also arrange basic workshops and training programs in these areas. However, there is a dearth of expert information professionals having competencies to successfully plan, manage, and initiate digitization projects, establish digitization labs, digital libraries, and online repositories.

*Terrorism threats and attacks* Pakistan has recently faced the worst terrorist attacks and

threats to the security of its educational institutions by extremists with political or religious ideologies. In the near past, burning of libraries by extremists in our neighboring country, Afghanistan, has been the worst damage done to the documentary heritage of mankind by eliminating important artifacts and books forever. Such factors are a constant threat for libraries and archives in Pakistan.

*Absence of legal control system* There is enough legal regime for governing information in Pakistan in the form of acts and ordinances to protect archives and rare materials. However, there is no governmental control system available to ensure their security, by listing or cataloging their locations and availability in public and private collections. We are aware of some valuable rare books in university professors' personal libraries, which after their death, were discarded by their families due to sheer ignorance about their value.

### **The Case of Nur-i-Afshan Archives**

The Ewing Memorial Library has very old and rare books and journals. A big bulk of valuable heritage of human knowledge had been carefully collected and keenly preserved by the past generations of librarians and caretakers. It is now in the hands of the librarians of the digital age, with a huge challenge of preservation and saving for the future generations.

The Library had seen some dark ages during the period when under a Governmental decision, the FCC was nationalized. The neglect and destruction caused by mishandling and lack of necessary everyday care of the collections like dusting and minor repairs, misuse, termites and mold had resulted into lost and torn papers and documents. Despite an obviously irreparable loss that had already happened, some part of the collection was still safe. A nearly complete back file of *Nur-Afshan*, a weekly Newspaper, was an interesting item among the old collection. *Nur-Afshan* is an almost extinct News Magazine of 18th Century which published initially from Lodhiana and afterwards from Ambala and Lahore during years 1873 to 1944. Its back-file

was found on the lowest shelves of the last row of collection on second floor, which is the hottest part of the Library and also and worst exposed to all environmental dangers. The dust-laden volumes were in a bad shape. Upon inspection, it appeared to be a multifaceted news magazine published weekly and bi-weekly. The contents looked very interesting with news summaries, both local and international; government postings; commodity prices; advertisements; opinion articles; essays; proverbs and poems – a wealth of information about the everyday lives and thought of people living three centuries ago. This was an interesting combination of public opinion of that time expressed in various writing formats, poetry, stories etc.

Interestingly, *Nur-Afshan* was not only the domain of professional journalists and academics, but also writings from the broad public. The magazine's editorial commitment to present contrasting perspectives was different in style and format as compared to the Urdu and English newspapers published during the same period. Its policy to accept unsolicited articles encouraged the portrayal of views from a variety of contributors also adding them to its archival importance.

*Nur-Afshan*, with its 18th Century classical Urdu language and printed in hand-written Urdu typeset, represented a multitude of voices discussing an array of social, political, and religious topics that demonstrated continuity with the past and, in later editions, the encounter with the modern. Certainly the magazine was a rich source of primary information – presenting mostly the happenings in the religious scenario in the sub-continent at that time. There was no doubt that the publication includes a variety of materials useful for researchers. It is one of the earliest Urdu publications and thus testifies to the evolution of Urdu usage in prose, news, and scholarly writing.

Articles, letters to editor and other information published in *Nur-Afshan* have references in a number of books published during that time and later, which clearly indicates the authenticity of

the primary information contained in it as well as an evidence of it being well read and popular among learned public of that time. Interestingly, we find the mention of its availability in FCC Library as follows:

“A complete set of *Nur-Afshan* volumes in Urdu is kept in the library of Forman Christian College, Lahore. The quotations in this book are taken from the English translation of the *Nur-Afshan* Letters of Sunder Singh, now available in the Central Library of Zurich.” (Appasamy, 1958 p.40).

Some volumes and issues were missing in the *Nur-Afshan* back file. We checked the availability of *Nur-Afshan* issues in local as well as regional libraries in the hope to find other copies of *Nur-Afshan*, and to borrow the missing pages to complete our back file. To our surprise, no other copy of *Nur-Afshan* was available anywhere else in Pakistan or in India. This meant whatever we had in our Library, was the last copy of this information available to the world – And the last chance for it to survive – only if we could digitize it!

To avoid further damage to the magazine it was decided to keep it under lock and key and to be consulted only after the permission of the Chief Librarian, and under careful supervision of a library staff member. It stayed as such for some years.

A demand for its more frequent usage aroused when a new Research Centre was established at Forman Christian College. The researchers required to consult *Nur-Afshan* for their studies more frequently and freely than our ‘Restricted usage policy’ could allow. A meeting with the Director was held to explain the situation of the material, demonstrating Library’s intention to keep it as safe as possible until its permanent preservation by digitization. Not only, the Director understood the issue but also introduced us to the *British Library Endangered Archives Program* as a possible source of grant money to support the digitization of *Nur-Afshan*.

### **The EAP660 – Project for Digitizing Nur-i-Afshan Archives. The British Library Endangered Archives Programme (EAP) aims to:**

Contribute to the preservation of archival material that is in danger of destruction, neglect or physical deterioration world-wide. This is achieved principally, through the award of grants in an annual competition. The grants provide funding to enable successful applicants to locate relevant endangered archival collections, to arrange their transfer to a suitable local archival home where possible, to create digital copies of the material and to deposit the copies with local institutions and the British Library.

In August 2012, the author prepared the estimated benchmarks for digitization workload, by looking at the total number of pages to be digitized, the specifications of scanner required keeping in view the size of pages and the physical condition of the magazine to be digitized, number of staff required for digitization, their qualifications and skills requirements and the estimated time period for the completion of the task. A budget was also worked out to estimate the grant requirement.

With this information in hand, a three members team started working on the preliminary grant proposal submitted to the British Library EAP in November 2012. After due assessment of the preliminary application the EAP administration invited us to submit a detailed application which was submitted in February 2013 and accepted for the 2013 award as a Major project, entitled, “*EAP660: Nur-i-Afshan archives: Perspectives on the inter-religious history of Punjab from 1873-1944*”. A grant of £35,600 was approved for this Project for a period of 12 months. Details of this Project are available on the British Library Endangered Archives Program Website.

### **Installations and Induction**

*The equipment* The Grant Money approved for the Project was received in October 2013. Keeping in view the condition of the Archive, the demands of the Endangered Archives Program, scanning standards and high quality imaging

requirements, a high-resolution overhead scanner was required. This type of scanner was not commonly available in Pakistan. Only one Company named “CANNON” supplied such a scanner in Pakistan and provided the after sales support and services. This fact was mentioned in the grant proposal and the funds had been already approved. However, the company required minimum 4 weeks prior confirmation of the order to import the machine. Upon the receipt of the grant money in November, 2013, the order for a CANNON Book Eye Scanner, two Computers, a Printer, UPS and digitization software was placed.

During the wait for receipt of the equipment, the project planning tasks were completed. Work procedures, work flow charts, policies and guidelines were developed for scanning staff. Also appropriate metadata schema was selected and controlled vocabularies for various metadata fields were prepared and the workflow instructions were documented. Finally, the job descriptions for imaging and scanning and metadata description staff, were defined for the hiring of two digitization staff members for the Project.

The Digitization Lab was prepared in a room and the CANON Book Eye4 scanner was installed and configured with a static IP address on Campus LAN. The staff work tables, computer workstations, Internet Nodes and appropriate Electricity installation completed the space preparation for the Project. The two Project staff members, a Metadata Librarian and a Scanning and Imaging Assistant were hired in February 2014. Given the situation of archives education in Pakistan, the skilled staff could not be expected. A well-rounded Cataloger was hired and trained for Metadata skills. Another with some imaging skills was selected for training on imaging standards, scanning software and data security and storage processes.

#### *Staff Training and Development*

There was another Project of British Library Endangered Archives Programme, entitled “EAP566: Endangered Urdu periodicals:

preservation and access for vulnerable scholarly resources”, which had started a year ago, and was fully functional in the Mushfiq Khawaja Library in Karachi. It was expected that the Project staff would have developed considerable experience of working according to the EAP guidelines and standards. A visit to Karachi was arranged in the last week of February 2014 to seek help for our project and to learn from them by observation and practical demonstration. The very helpful staff of EAP566 Project happily and generously shared their experiences with us, answered our questions, demonstrated the scanning process and metadata listing sheets and provided very useful tips for imaging quality. We learned much more by practical observation of the digitization process according to the British Library standards than we could by reading the instructions and guidelines.

#### **The Bottlenecks and Barriers**

##### *The Scanning Blues*

*The KODAK color patch card.* The project was ready to start on the very day when the first bottleneck appeared. EAP guidelines for imaging and scanning, essentially required us to produce the scanned images output of Nur Afshan pages, in “uncompressed TIFF format, or in uncompressed RAW format”, using “an industry-approved colour chart or greyscale reference card”, with each page image. This was required for the EAP staff to check the accuracy of the color captured in the photographs. This technique helps in the post-production processes of the scanned images before uploading them to the online repository.

The color chart or color checker is a small item of 8-inch or 20 cm length and costing around 20-25 dollars, which is normally available in common photography and camera shops. However, to our disappointment, it was not available in Pakistan. We tried to place an order with the photography shops, but being a low-cost item, no one was ready to accept an import order. After many failed efforts, we explored the help from our good old friends – the booksellers. Luckily, one of them agreed to import two sets of Kodak

Color Separation Guide and Gray Scale (Small), Q-13 Color control patches.

*Document cleaning tools and supplies.* The archival supplies for which we see elaborate archival supply websites and stores offering a wide variety of supplies for each minor activity in digitization process, is just a dream in Pakistan. We came to know that even minor digitization supplies like document cleaning brushes, document-holding weights, image flattening glass, cleaning accessories and other everyday use tools and equipment for the Digitization Lab, were not available in Pakistan. We had to import small items and waited for long time in the import process. In some cases, small and low cost items could not be imported. In such cases, we explored for some funny indigenous solutions. For example, the document cleaning brushes, a necessity for digitization, were replaced by common cosmetics brushes from beauty shops.

*The Malinax sheet.* There was a need for 3 X 2 ft. pieces of scratch proof and glare proof glass or Malinax sheets to flatten the page images for the purpose of clearer and readable scanned images. This could neither be found locally, nor could be imported because no supplier was ready to import such a small quantity of a specialized item. Having no other option, we decided to use normal windowpane plastic sheets. Soon it was found that the electro-static charge on the sub-standards plastic sheet was picking up large amounts of the loose particles of paper on the decaying while causing very fine scratches, which resulted into the blurring of the surface. These sheets get scratched easily. After changing sheets frequently, I checked with the shop to check what material it was. I was told it was Acrylic which reminded me the Acrylic paint of my car. Next day, I picked the car polish tin from home and asked the staff to buff the sheets. The result was a scratch-free and less static sheet. This added a new Lab routine to buff the window plastic sheets weekly with scratch removing car polish. We used them in Dark Room conditions to avoid glare during scanning.

### **The First Success! Hurray!**

Meanwhile, we downloaded a printable Color Patch Card from Internet and printed on photo paper in high resolution print and started producing sample scanned images. After some initial testing and trials, the digitization work formally started on March 13, 2014. As an EAP instruction, the first set of Sample Images and Metadata Listing has to be uploaded on FTP Directory of EAP for checking by the EAP Curator to ensure that the work done is actually in compliance with the British Library Standards and guidelines. This data was uploaded on March 24, 2014, for inspection and approval and / or feedback by the EAP staff. On April 7, 2014, Jody Butterworth, Curator, EAP responded as follows:

Thank you for uploading the material to the ftp account. I have managed to have a look at the images today and I have to say they look extremely professional and I believe they comply to the EAP guidelines. However, I would just like to ask about the colour checker, you had mentioned that you had been unable to obtain one, am I right in thinking therefore that the checker in the images is a photocopy reproduction? Of course ideally a real colour checker should be used. Do let me know if it is a reproduction. (Personal Communication).

We reported back that it was actually a printed one and not a photocopy. However, we had to stop further scanning until the end of May, 2014 in wait of receipt of our import order for the color checker. The regular scanning started on March 13, 2014 and completed on March 13, 2015. The Project scanned a total 31401 Pages and produced 3-Terrabytes of scanned data. None of the express mail companies accepted the delivery of computerized data to UK. So, the three Terrabyte Hard Disks were hand carried by an FCC scholar traveling to UK. The few remaining volumes were uploaded on the FTP site on to the Endangered Archives Program on May 14, 2015.

The Complete Digital Collection of Nu Afshan Archives was made Online, on June 2, 2015. A



special Blog Post announced its addition as a special Project due to which, EAP crossed a landmark of having 5 million images. (Endangered Archives Blog, 2015) The last copy of 18th Century dying and endangered periodical has been saved in British Library Repository. It is now universally available online for the use of students, historians, researchers and readers at large. Endangered Archives (2013).

### **Managing the Online Access and Retrieval**

*Need for an online repository.* According to Terms and Conditions of EAP grant, the EAP660 was required to do the following activities:

1. Upon completion, the EAP660 Project will supply the high-resolution digitized archives of Nur Afshan produced under an EAP grant, for depositing to the Programme's Archive on the British Library website.
2. In addition to that, copies of digital scans produced under the grant will also be deposited in a local repository.
3. In general, EAP priorities for digital copies of material are that they should be permanently available, useful for the recovery of knowledge, and universally accessible.

No local digital repository existed in the country to share the deposit of the digital images of EAP550, to fulfill the grant conditions of keeping them permanently available and universally accessible for research purposes.

There was a need to develop an online digital repository for this purpose. The knowledge and skills for developing an online digital library system were available, however the hi-volume and hi-speed storage, retrieval networking infrastructure, high-speed access and retrieval software infrastructure were very expensive and beyond the budgetary resources of both the EAP660 project as well as its Archival partner, the Ewing Memorial Library.

Hence the support of a third partner was definitely required and luckily, it was available.

*The AMICAL consortium* Ewing Memorial is a member of the AMICAL- an International Consortium of American-model, liberal arts institutions of higher learning. We also had subscribed to OCLC RESPOND which is a joint AMICAL-OCLC initiative for AMICAL Members.

Apart from offering participation in Worldcat Union Catalog and Worldshare Resource-sharing Programs, the OCLC also provides a free hosting instance of CONTENTdm – the Digital Collection Management Software of OCLC, to AMICAL members subscribing to the RESPOND Program. We decided to use this facility to build the Ewing Memorial Library Digital Archives.

### **Nur-i-Afshan Online Digital Archives**

An online access authorization to the CONTENT dm Digital Library Software was obtained and the digital storage site for the Ewing Memorial Library Digital Archives was created and configured on the OCLC Cloud Server.

Then the online archive collection, the *Nur Afshan Digital Archives*, was created and configured for uploading the images produced by the EAP660 Project.

CONTENT dm software offers a number of standard metadata schemas to select an appropriate one according to the type of items as well as the subject matter of the digital repository. Dublin core metadata format looked more appropriate in terms of fields available for description, so it was selected for building a repository structure for Nur Afshan Digital Archives. By now, a test version of the Archives has been published where it can be viewed anytime through Internet. The success of this initiative is the rare example of an unusual strong collaboration and successful pooling and sharing of resources available from multiple partners:

- The British Library Endangered Archives Program
- The Forman Christian College
- The Ewing Memorial Library
- The AMICAL Consortium
- The OCLC – RESPOND Program

### Major Issues of the Online Local Digital Library

*The Urdu OCR.* The basic language of Nur-i-Afshan Archives is Urdu. The users of the British Library cannot read or understand Urdu language. So, uploading the repository with the static non-editable images of the Newspaper Pages with the metadata described in English language or Romanized English was an acceptable solution. However, in the local repository, we could have easily offered full-text searches and Urdu Metadata searches for the users who are well versed with the Urdu Language.

But, unluckily, there was no facility for an Urdu OCR available in Pakistan. While searching for literature, a number of evidences of use of Hindi, Marathi, Tamil, and Sanskrit OCR (Bag & Harit, 2013) and Bangla and Devanagari scripts OCR Software (Indsenz, 2015) were found in neighboring countries, India and Bangladesh. Yet Pakistan is lagging far behind in this important area of Digitization.

*Controlled vocabulary and indexing language.* Nur-i-Afshan newspaper covers a variety of thematic areas specific to the India-Pakistan subcontinent. The period covered by the archive is particularly important in the sense that many political and religious movements and events for which a comprehensive thesaurus of Religious and political terminology was required.

No comprehensive thesaurus of terms could be found in Urdu or English. Finding appropriate terms for defining metadata for the Dublin Core Indexing fields was a big challenge. The whole archive was in old classical Urdu language of 18th Century India. Reading and understanding the concepts in Urdu terminology and then finding appropriate terms in English, was a really challenging and difficult task. A number of online catalogs and in- print glossaries, dictionaries and thesauri had to be consulted to build a list of terms to be used to define the subject content of the archives.

*The Digital Surrogate vs. Original Documents:* With the completion of the EAP 660 Project the

targets of preservation of an information resource on dying paper have been successfully achieved. But the questions, that weather the digital surrogates of Nur-i-Afshan, will have the same value of being a historical evidence as the original newspaper back file documents do, still remain unanswered. We can only try our best to recreate the original document in as accurate color and look of the original document as possible. But the images produced digital format still remains a surrogate and not the original. Issues of originality, quality, accuracy and truthful representation will still remain. With the proliferation of digital images editing and image manipulation tools, ensuring the security and originality of the images and digital surrogates, will be a constant challenge. The original documents have lived their life – But we are still faced to keep the originals or discard them.

### Conclusion

1. The University libraries in Pakistan together hold enormous number of archival information resources in analog print formats which are deteriorating and destroying with time. Digitization of these resources is an essential process to publish them online and facilitate remote access to users.
2. There is a scarcity of literature regarding digitization initiatives in Pakistani libraries. However, some reported studies have revealed that the pages of the documents are being scanned using a scanner to create the digital images, without any care for Digital Imaging Standards of Storage and Display or keeping in view the image quality for preservation and reuse. The scanned images are mostly being stored on the CDs and DVD-RoMs or Computer hard disks and this is considered as the whole of digitization process by the librarians. Such images are constantly in danger as the CD-ROM is not a reliable storage medium for archival purposes. Even computer hard disks are not reliable for permanent storage. Sustainability of such efforts cannot be

ensured in terms of storage media, scalability of storage formats and usability for online storage and retrieval in future.

3. Such digitization work done without the assurance of compatibility with International standards, actually carries no more value than the digital
4. Replicas of print collections on the shelves. The digitized documents in this case, may only be accessed and retrieved physically by visiting the library and may only be physically delivered to the users through a hand-carried storage medium. The document images in such cases can be viewed and browsed serially one by one within the folders or the portable document files (PDF).
5. Online searching, remote access and on-screen display before retrieval is not possible. No care is actually considered necessary for the scanner specifications to create archival quality images in compliance with the preservation standards for scanning. Assigning proper metadata to the images, for online searching and retrieval, is also not in practice.
6. There is a danger of loss of all the efforts of scanning, if at the time of building an online repository in future, the quality and types of formats of digital images, does not comply with the standards of online storage and display.
7. In the absence of proper metadata descriptions, the thousands of scanned images stored on the computer disks will be just like huge collections of un-cataloged books in the library. Like the retrospective conversion of print catalogs for implementing library automation projects was a big cause of delay in the past, the stored images will have to go through a long process of metadata description, before a sustainable online search and retrieval system may be installed.

### Recommendations

1. The indigenous information resources located in university libraries of Pakistan are our collective national assets. Lack of knowledge and expertise and inability to plan successful digitization projects should not be allowed to damage these assets.
2. There should be a Master Degree specialization for Archives Management and Digitization at the Universities to fill the gap of education and expertise. The National Standards and guidelines should be developed to ensure standards, security and interchangeability of the scanned archives
3. Thesauri should be developed in local languages, indigenous terminologies and concepts to capture the local knowledge in standard terminology.
4. Digital Repository Software and OCR Software in local languages should be developed to digitize local information and to provide search and retrieval services.
5. The libraries interested in archival digitization in Pakistan, should organize themselves in strong networks or consortia and should plan collaborative projects by pooling up their resource and cooperate to learn from each other. The consortia may start formal trainings for their staff to develop the required skills and to work together to formulate standards, policies and guidelines for sustainable projects of archival digitization.

### References

- Ahsan, M. (2009). Digital divide and digitization initiatives in Pakistan: a bird's eye view. Printed on demand by "Nuova Cultura" www. nuovacultura.it Retrieved on 3.10.2015 from [http://elpub.scix.net/data/works/att/78\\_elpub2009.content.pdf](http://elpub.scix.net/data/works/att/78_elpub2009.content.pdf)
- Appasamy, A. J. (1958). *Sundar Singh: A Biography*: Christian Literature Society.

- Bag, S., & Harit, G. (2013). A survey on optical character recognition for Bangla and Devanagari scripts. *Sadhana*, 38(1), 133-168. 14.
- Endangered Archives (2013). "Nur-i-Afshan archives: Perspectives on the inter-religious history of Punjab from 1873-1944." Retrieved 27-9, 2014, from [http://eap.bl.uk/database/overview\\_project.a4d?projID=EAP660;r=41](http://eap.bl.uk/database/overview_project.a4d?projID=EAP660;r=41).
- Endangered Archives (2015). "Endangered Archives Program (EAP)." Retrieved 3-10, 2015, from [http://eap.bl.uk/database/overview\\_project.a4d?projID=EAP660;r=41](http://eap.bl.uk/database/overview_project.a4d?projID=EAP660;r=41). & [http://eap.bl.uk/database/overview\\_project.a4d?projID=EAP566;r=41](http://eap.bl.uk/database/overview_project.a4d?projID=EAP566;r=41)
- Endangered Archives Blog (2015). "'Voices from pre-partition India' - The Nur-i-Afshan periodical from the Punjab." Retrieved Oct., 2015, from [http://britishlibrary.typepad.co.uk/endangeredarchives/2015/06/urdu\\_periodical\\_punjab.html](http://britishlibrary.typepad.co.uk/endangeredarchives/2015/06/urdu_periodical_punjab.html).
- Forman Christian College (n.d.). "Heritage." Retrieved 3-10, 2015, from <http://www.fccollege.edu.pk/about/heritage>.
- Indsenz. Accurate and fast digitization of Hindi, Marathi, Tamil, and Sanskrit. <http://www.indsenz.com/int/index.php> accessed dated 3.10.2015
- Ghayyur Hussain, Syed (2012). Digital initiatives by the department of libraries: case study of automation in the National Library of Pakistan, PLA Newsletter (Special Issue, Election 2012). 12-20.
- Indsenz (n.d.). "Making sense of Indian documents: accurate and fast digitization of Hindi, Marathi, Tamil, and Sanskrit." Retrieved 3-10, 2015, from <http://www.indsenz.com/int/index.php>.
- Iqbal Academy Pakistan (2014). "Allama Iqbal Cyber Library." Retrieved 3.10.2015 from <http://www.iqbalcyberlibrary.net/en/>.
- Sindh Archives (2013). "Islamic Manuscripts in Pakistan." 3.10.2015, from <http://www.sindharchives.gov.pk/articles.aspx>.
- UNESCO. Memory of the World Program. Retrieved on Date 2019/09/21. <http://www.unesco.org/new/en/communication-and-information/flagship-project-activities/memory-of-the-world/homepage/>,
- Verzosa, F. A. M. (2004). Digital initiatives in archival preservation. International Conference on Challenges in Preserving and Managing Cultural Heritage Resources. held on October 19-21, 2005 at the Institute of Social Order, Ateneo de Manila University, Loyola Heights, Quezon City, (Philippines). [https://paarl.wikispaces.com/file/view/Digital\\_Initiatives\\_in\\_Archival\\_Preservation.pdf](https://paarl.wikispaces.com/file/view/Digital_Initiatives_in_Archival_Preservation.pdf) Retrieved 3.10.2015
- Rafiq, M. & Ameen, K. (2013). Digitization in university libraries of Pakistan. OCLC Systems & Services, 29(1). 37-46.