

# "Digital Initiatives in Archival Preservation"

by

Fe Angela M. Verzosa

## **ABSTRACT:**

For digital archives, the innovative use of information technology will enable archival information to be accessed anywhere anytime by anyone who needs them. This is the ultimate objective of archival preservation---to preserve materials of enduring value for the sake of posterity. However, the need to integrate technology and content is evidently significant. Funding assistance and the commitment of archivists and information professionals in such digital initiatives are critical to its success.

## **ABOUT THE AUTHOR:**

Fe Angela M. Verzosa is the University Archivist of De La Salle University-Manila since 1988, and served as its Library Director from June 1994 until May 1998. She graduated with an A.B. History degree (1966) and a Master in Library Science (1978) from the University of the Philippines, Diliman. She served as Head of the National Committee for Culture and the Arts (NCCA) - National Committee on Libraries and Information Services from 2001-2004. She was President of the Philippine Association of Academic and Research Libraries (for two terms, 2000 and 2004) and past President (2000) of the Philippine Librarians Association, Inc., an accredited professional organization of licensed librarians.

# "Digital Initiatives in Archival Preservation"\*

## Introduction

First, I wish to express my appreciation to the organizers for inviting me to speak on a topic that is close to my heart, that is, "Digital initiatives in archival preservation."

This presentation will discuss how, with the increasing use of Internet and electronic resources, digital archives are now an emerging platform for cooperative preservation efforts. This presentation will also explore local digital initiatives here and through the Internet to view the different frameworks and models. Finally, it will recommend digital technology as a means towards preservation of original archival materials.

Now, let's look into some definitions before we proceed with the topic at hand...

"Digital preservation" or "digital archiving" means taking steps to ensure the longevity of electronic documents. It applies to documents that are either "born digital" and stored on-line (or on CD-ROM, diskettes or other physical carriers) or to the products of analog-to-digital conversion, if long-term access is intended.

Now what do we mean by "Digital archives" ? These refer to Archival materials that have been converted to machine-readable format, usually for the sake of preservation or to make them more accessible to users. They may also refer to information originally created in electronic format, preserved for its archival value.

## Digitization

Let me start by making a sweeping statement: **Digitization greatly enhances access, faster and easier.** The big idea is to digitize often hidden or difficult to access archival collections. It allows remote access (via Internet/cd networking) by simultaneous users. As an activity, it generates publicity for institutions because it gives them the opportunity to showcase their rare or unique collections online. If digitized materials serve as surrogates to the originals, digitization can enhance preservation efforts by reducing the handling of original documents. Reduced access to the original helps to protect them from loss or mutilation, etc. The digital copy may serve also as backup in case of disaster, like fire, flooding, etc.

---

\* Paper presented at the 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.

Digitization is cheaper than photocopying in terms of low cost distribution of huge data. It also provides reduced long-term storage costs and greatly reduces document storage space by 80%.

Digital preservation, furthermore, has an inherent appeal, particularly for administrators, patrons, and funding agencies. It is an activity well supported because it provides mass sharing capabilities and can generate a lot of publicity.

However, one of the growing concerns in this endeavor is that there is a growing amount of born digital and born-again digital material, which are intended as substitute to the original materials. Can digitization serve a preservation function? Yes, if the primary goal of digital archives is long-term preservation of information in a form that guarantees sustained access and dissemination in the future. Can digitization fulfill this goal? This is the main focus of our discussion.

## Digital preservation

This issue is a high debate topic, and the longevity of electronic records, the increasing obsolescence of computer hardware and software, and the storage problem for digital media, or the fragility of digital media (its 'shelf life' compared with, say, non-acidic paper), remain a huge problem. Along with technological advances is the responsibility to develop a continuing program to assure access to preserved information over time.

There are more problems, like :

- Getting the information into digital form** (providing index fields, "grooming" – data entry and cleaning, etc. ; issues like will scanning further damage deteriorated originals? Automatic sheet feeding may cause damage to brittle materials)
- Keeping it in digital form** (backing up, refreshing the digital files, and migrating them; upgrading hardware and software over time)
- Standards to ensure a high quality result** (high quality scanning process and digital media)
- Copyright/intellectual property implications**
- Maintaining the integrity of digitized information over time** (maintaining server security and ensuring that restoration of applications and data from backups is always possible) Digital files reside on unstable media that require frequent migration of data. Ensuring the integrity of the data is significant in this repeated process.

According to Margaret Hedstrom, of the School of Information at the University of Michigan (Ann Arbor), " The difficulty and expense of preserving digital information

is a potential impediment to digital archives and library development. Preservation of traditional materials became more successful and systematic after libraries and archives integrated preservation into overall planning and resource allocation. Digital preservation is largely experimental and replete with the risks associated with untested methods.”

## Digital preservation - issues and concerns

There are other issues and concerns in digital preservation that may be summarized as follows:

\*Digital preservation strategies are shaped by the needs and constraints of repositories with little consideration for the requirements of current and future users of digital scholarly resources.

\*Aside from the costs of maintaining digital data, the cost of preservation is separate from the cost of access.

\*What form to keep digital data in??

This element may include the cost of purchase or design of any software or hardware needed to prepare an object for archiving. What is required to determine the best technical strategy for providing continuing access to material in the archive – i.e. migration, emulation. Migration is the process of transferring data from a platform that is in danger of becoming obsolete to a current platform. This process has both dangers and costs. Emulation is an alternative and superior strategy. The essential idea behind emulation is to be able to access or run original data/software on a new/current platform by running software on the new/current platform that emulates the original platform. This will be determined by agreement on an object’s significant properties. For example, an object which requires preservation of its "look and feel" may require the development or enhancement of emulation tools.

\*Authenticity - securing digital objects against unauthorized changes

\*Technological Obsolescence

\*Additional information we need to keep (metadata)

Some metadata may be collected or incorporated from existing cataloging or other metadata records

## More concerns in digital preservation ...

For pictures, sound, and other archival multimedia materials which are storage-intensive, the opportunities for cooperation by establishing as many websites as possible offers a better alternative than a single site. A single site will definitely be burdened with a big storage requirement, while building separate sites will encourage cooperation. The

more participants involved, the better for the project to achieve efficiency and effectiveness.

For digital archives, the innovative use of information technology will enable archival information to be accessed anywhere anytime by anyone who needs them. However, the need to integrate technology and content is evident. And outside funding and the commitment of volunteers in such digital initiatives are critical to its success.

Digitization, however, must not be viewed as a substitute for other preservation activities. Microfilming remains the preferred long-term preservation method in this age of technology. Preserving the original materials increases the intrinsic value of the original as evidence.

Unquestionably, the Internet presents one of the most effective means towards making access to archival materials available through networking. However, I believe we must recognize that the value and focus of the Internet is full text rather than a mere catalog of bibliographic information. At this juncture, there is a more urgent need for content on the Internet, and this is something that we need to carefully explore as the focus of cooperative projects.

Let me now give you a few models and frameworks for some early digital initiatives.....

## **Five Colleges, Incorporated**

This is a nonprofit educational consortium established in 1965 to promote the broad educational and cultural objectives of its member institutions, which include four private, liberal arts colleges and the Amherst campus of the state university. The consortium is an outgrowth of a highly successful collaboration in the 1950s among Amherst College, Mount Holyoke College, Smith College, and the University of Massachusetts Amherst, which resulted in the founding of a fifth institution, Hampshire College, in 1970.

It developed a Web site which provides access to digitized versions of archival records and manuscript collections relating primarily to women's history -- particularly women's education at the Five Colleges. Included among the collections are official college publications, letters, photographs, articles, oral histories, diaries, and more. .

The completed project encompasses 54 online collections, comprising over 38,000 items. These collections are accessible through the Web site, through electronic finding aids provided on the Web sites of the repositories at each institution, and increasingly through records in the Five College online catalogs.

## C e d a r s Curl Exemplars in digital archives

Cedars began as a project of CURL (Consortium of University Research Libraries in UK and Ireland) in April 1998 and ended in March 2002. Its broad objective was to explore digital preservation issues. These range through acquiring digital objects, their long-term retention, sufficient description, and eventual access.

As part of the Cedars Project Demonstrator Archive, the Cedars Project has included a number of "test sites" designed to test different aspects of digital archiving with various types of digital materials (image or text files, electronic journals, large online databases and other complex multi-media materials). The Project chose organizations with a breadth of expertise and interests and has worked with each one to explore different aspects of digital archiving best suited to their situation and interests. Based on these discussions with the sites, the Project designed a tailored web-site area where the test sites could use the Cedars Demonstrator Archive and evaluate its effectiveness.

Test sites include three universities in the United Kingdom (Birmingham, Exeter, and University College London), the British Library, and Birmingham Central Library. Primary funding provided by the Joint Information Systems Committee (JISC) through its eLib and Digital Preservation Focus programmes.

## Charette Digital Project

A model framework for a digital initiative is the Pittsburgh Architectural Archives' Charette Digital Project of Carnegie Mellon University. This project has created an Online access to *Charette* magazine (vols. 1-54, 1920-1974), an essential source for architectural research in Pittsburgh and Pennsylvania.

Electronic images of each page include all text, photographs, graphics, advertisements, etc. The resulting database permits browsing by issue and full-text searching of the contents of *Charette*. There were five primary stages to this project:

- Clarify copyright
- Identify and obtain all issues of Charette
- Scanning to be done by Digital Library Initiatives
- Utilize journal delivery and navigation functionality using the DIVA system (CMU's library system)
- Develop web interface

## USGenWeb Project

USGenWeb Project is managed by a group of volunteers working together to provide Internet websites for genealogical research in every county and every state of the United States. This Project is non-commercial and fully committed to free access for everyone.

The USGenWeb Digital Library (Archives) was developed to present actual transcriptions of public domain records on the Internet. This huge undertaking is the cooperative effort of volunteers who either have electronically formatted files on census records, marriage bonds, wills, and other public documents, or are willing to transcribe this information to contribute.

The USGenWeb digital map library is another outgrowth of The USGenWeb Archives. The United States Digital Map Library is a new project currently being developed by USGenWeb Archives. The goal is to make available to genealogists, useful, readable, high quality maps. Here you will find both archival maps and newly made maps based on scholarly research.

## Online Archive of California

A core component of the California Digital Library, the Online Archive of California (OAC) is a digital information resource that facilitates and provides access to materials such as manuscripts, photographs, and works of art held in libraries, museums, archives, and other institutions across California.

The OAC is available to a broad spectrum of users -students, teachers, and researchers of all levels. Through the OAC, all have access to information previously available only to scholars who traveled to collection sites.

### Collection Highlights:

- Primary sources include letters, diaries, manuscripts, legal and financial records, photographs, maps, architectural and engineering records, artwork, scientific logbooks, electronic records, sound recordings, oral histories artifacts and ephemera.
- Over 120,000 images; 50,000 pages of documents, letters, and oral histories; and 8,000 guides to collections are available.
- Offering easy-to-use search and viewing tools, the OAC organizes images into thematic and institutional collections, such as historical topics, nature, places, and technology.

## Digital Projects - Library Gateway

The University of Illinois – Urbana at Champaign (UIUC) Library and its departmental libraries conduct a broad range of digital projects, including the digitization of library collections and archives, research in information retrieval technologies for digital collections, and digital library research projects. They also develop instructional resources and facilitating tools for digitization projects, and experiment with the application of digital content in teaching and learning.

Among its projects are:...

- ALA Archives Digital Collections - includes photographs of past ALA conferences and images of library buildings around the country
- Art Museum Image Consortium - contains over 100,000 works of art, including paintings, sculptures, drawings and watercolors, photographs, and more available only to those with library accounts
- Digital Cultural Heritage Community - Materials from East Central Illinois local area museums, archives and libraries
- Global Cultural memory Project - Integrates materials from collaborating institutions, focusing on the past fifty years of cultural history in Champaign County, Ill.

### Digital Gateway to Cultural Heritage Materials

Of the above-mentioned projects of Library Gateway, the Digital Gateway to Cultural Heritage Materials is one of the most important projects of the University of Illinois (UIUC). This project:

- Contains approximately 538,485 metadata records from 32 collections of images, text, sheet music, websites, museums and archives.
- Includes the Alex Catalogue of Electronic Texts, a collection of public domain documents from American and English literature as well as Western philosophy.
- Includes 2820 scanned maps representing rare 18th and 19th century North and South American cartographic history from the David Rumsey map collection
- Includes 3127 images of coins and other artifacts from this institution, which advances the appreciation of coins, medals, and related objects of all cultures

### McGill University

The Digital Collections Program (DCP) of McGill University has been created to digitally capture and provide electronic access to McGill University's unique rare and special collections via the World Wide Web using advanced digital technologies. These collections include rare books, maps, manuscripts, prints, photographs, sound recordings, and more.

To date since 1997, DCP has produced more than forty digital projects covering a wide array of subjects including, art, architecture, history and literature, engineering, medicine, maps, music, and urban design. The collections range from archival inventories to bibliographic databases which provide access to items in the collection; image databases, and full-text collections



## JACAR

Here in Asia, Japan and China lead in digital preservation initiatives. Internet users can read a large amount of Asian historical records from the Meiji Restoration in 1868 until the end of the WWII in 1945, download and print it out easily. About six and a half million images have been scanned, and around two million images every year will be added. All the documents were previously microfilmed and converted into digital format

Spearheaded by the Japan Center for Asian Historical Records (JACAR), a full-fledged digital archives now provide image data of official documents concerning the relationship between Japan and its Asian neighbors through the Internet. It offers "anyone, anytime, anywhere" access to the records free of charge.

## JARDA

Another model of collaboration is JARDA, the Japanese American Relocation Digital Archive, which results in a single point of access to digitized archival materials on the evacuation and internment of thousands of Japanese Americans during WWII. Curatorial and digital collaboration was made possible through California Digital Library's Online Archive of California, which spearheaded the development of digital archives for "thematic collections." Putting all these primary documents in one site made a big contribution to academic research.

## Inter-Institutional Consortium Project

In the Philippines, our early efforts at digitization began in the late '90s. Using state-of-the-art digitally programmed and engineered CD-ROM mastering and production, the Inter-Institutional Consortium, now known as the South Manila Consortium, developed a digital project for selected Filipiniana rare books to be stored in compact discs, with multiple copies made available to every member-library for the regular use of their respective clientele. This project was funded by the United Board for Higher Education and benefited 5 institutions (De La Salle University, Philippine Christian University, Philippine Normal University, St. Scholastica's College, and St. Paul University-Manila). The digitization project was continued by De La Salle University on its own initiative.

## De La Salle University Archives

De La Salle University Archives has digitized its theses collection as early as in 1998. To date, we have 460 doctoral dissertation in 42 cds, 2,443 master's theses in 92 cds, and 518 born-digital theses received from the Registrar's office. Its yearbook collection from the 1920s to 1961 were digitized, and from 1998 onwards, *Green and White* came in both print and cd versions.

The Archives also recently put on trial a few articles online under the journal title, *Sinupang Lasalyano*. This journal provides open access to all of its contents on the principle that making research freely available to the public supports a greater global exchange of knowledge, and such access is associated with increased readership and increased citation of an author's work. The contents are indexed and searchable by OAI engines. This approach uses an open journal system, designed by the Public Knowledge Project, to improve the scholarly and public quality of research, and the software is freely distributed to support the open access publishing of scholarly resources.

## National Archives

Known as the richest in Asia with 60 million documents, our National Archives, known now as the Records Management and Archives Office (RMAO), has started its project on digitization of archival documents since 1998 and is still on-going. More than 13 million historic documents have been digitized but not accessible in the world wide web. How many more such digital projects are ongoing in the country but are not accessible in the web, no one knows. Why? Clearly, there are problems in funding initiatives, technology infrastructure, and technical expertise.

Preservation efforts at The National Archives remain focused on microfilming. There was a joint microfilming project of our Records Management and Archives Office (RMAO) and the Consejo Superior de Investigaciones Cientificas (CSIC) on the initiative of our government and the Spanish Embassy, resulting in a turn-over of the catalogues produced by the CSIC of microfilmed centuries-old documents during the Spanish colonial period of my country. It is hoped that by using the model of a hybrid conversion, the institution can combine the advantages of scanning for access purposes with the preservation benefits of microfilming.

## UP Diliman Libraries

The UP Diliman Libraries began digital projects in the year 2000 with the Cumulative Index to Philippine Periodicals 1995-2000 in CD-Rom. The Harry Whitfield Harnish Collection (1898-1907) was another important project, funded by NCCA. This is a collection of photo artworks by an American soldier who retired in the Philippines after the war. The collection includes essays on issues covered by the photos. The Board of Regents' Minutes of meetings from 1925 to present were also digitized. Filipiniana, UP publications, researches, and other Special Materials in the UP Library System were also digitized as UP's contribution to the Philippine eLibrary Project, which I will discuss shortly. An ongoing project is the digitization of UP theses and dissertation.

## Philippine Library Network

The Philippine Library Network (known as the eLib Project) was recently launched in Sept 2004. This is a government-sponsored collaborative undertaking of major government libraries, where the core collections belong to The National Library,

the Department of Science and Technology (DOST), the Department of Agriculture, and the University of the Philippines Library System.

It envisions making available in digital format the holdings of The National Library (over 1.2 million volumes, including materials considered part of the national heritage), the UP Library System (over one million also), and the holdings of the SciNet libraries of the DOST, and the library networks of the Department of Agriculture. A total of twenty-four million images will be accessible in the world wide web ([www.elib.gov.ph](http://www.elib.gov.ph)).

While the focus is in making available Philippine materials on a global scale, the project also provides cooperative acquisition and sharing of expensive on-line subscriptions. Its data center, which will house its portal and database mirrors, will be conveniently located at The National Library, with a data recovery site at UP Diliman. The project will use open source and open-standard tools, which can be replicated by other libraries. For its financial sustainability, the project encourages institutional and individual subscribers to gain access to its specialized databases, and the management of its fee-based services will be administered by the UP Library System.

## Supreme Court Library

Digitization of the Supreme Court Library collection began in April 2004 with the World Bank loan and when the Library and Printing Committee was reactivated with Justice Antonio T. Carpio as Chairman and Justice Conchita C. Morales as Co-Chair. The project -

- Aims to be primary research in the speedy delivery of justice
- Serves the entire Judiciary - the Supreme Court, Court of Appeals, Sandiganbayan, Court of Tax Appeals, Regional Trial Courts, Metropolitan Trial Courts, Municipals Trial Courts.
- Contains:
  - SC Decisions and Resolutions
  - Rules of Court
  - Administrative Matters
  - Circulars and Orders

The memorabilia section is devoted to Chief Justices and retired Justices of the Supreme Court. It contains the picture, a brief biography, decisions, and available articles, and speeches of each retired Justice of the Supreme Court

## Summary

These examples lead us to the conclusion, rightly or wrongly, that developed nations can manage their own websites for digital archives without need of collaboration. Collaboration becomes an attractive alternative only for the poorer nations when better endowed countries take on the initiative to extend assistance by way of providing the

required funds, or the needed technical expertise and training, and/or the technology infrastructure or network mechanism that will facilitate access.

Here, our national government agencies in collaboration with private institutions, can take the initiative in creating digital networks of institutional repositories with rare heritage collections, matching institutions with common interests, needs and requirements.

## Copyright Concerns

One of the most serious barriers to building digital archives is related to copyright and intellectual property concerns. With these legal issues, it may be difficult to develop higher levels of cooperative arrangements, but the problem is not insurmountable. One of the potential solutions is to require waivers from intellectual property right-holders to the digital contents of their works in the spirit of shared scholarship and complementary research.

Certainly, many academic and research institutions can be encouraged to require their waivers, provided the collaborative partners are also willing to do the same and share their valuable resources. At present this willingness is still doubtful, and efforts towards digital collaboration have been focused on historical materials and government works free of copyright concerns.

## Conclusion

In conclusion, allow me to state here that now is the time to initiate collaborative programs for the future access to preserved digital and born-again digital materials for the sake of posterity. Thanks to digitalization, collaborative programs now extend far beyond the traditional modes of preservation. Equity in terms of cooperation should not pose any stumbling blocks because every archival institution has its own unique collection that can be offered to the international community of archive users, and in the context of free or almost-free.

As long as the primary goal of digital archives is long-term preservation of information in a form that guarantees sustained access and dissemination in the future, it will have a purpose in archival preservation.

The End

## REFERENCES

- Bertot, John Carlo (Associate Professor, School of Information Studies, Florida State University), *Building a (Maryland) Digital Library* (a powerpoint presentation, Nov 1, 2002)
- Bunker, Geri (Coordinator, Digital Library Initiatives, and Interim Associate Director of Libraries for Resources and Collection Management Services), “*Collaboration as a Key to Digital Library Development: High Performance Image Management at the University of Washington*,” *D-Lib Magazine*, March 1999 (v.5, no. 3)
- Hedstrom, Margaret, “*Digital Preservation: A Time Bomb for Digital Libraries*,” in [http://institute21.stanford.edu/Current\\_Programs/2000\\_Institute/Presenters/margarethedstrom.html](http://institute21.stanford.edu/Current_Programs/2000_Institute/Presenters/margarethedstrom.html)
- Heun, Christopher T., Courts, “*Unlikely To Stop Google Book Copying*,” *InternetWeek*, 4p.
- Lee Kyong-Ho, Oliver Slattery, Richang Lu, Xiao Tang, and Victor McCrary “*The State of the Art and Practice in Digital Preservation*,” *Journal of Research of the National Institute of Standards and Technology*, (vol. 107, no. 1), January–February 2002
- Matson, Lisa Dallape and David J. Bonski, “*Do Digital Libraries Need Librarians? An Experiential Dialog*,” *ONLINE*, November 1997
- Missingham, Roxanne (Resource Sharing Branch, National Library of Australia), “*Creating digital futures together: collaboration between the National library and Australian libraries*,” 6p.
- Steele, Colin (University Librarian, Australian National University) and . Mechthild Guha (Manager, Asia Pacific Cluster, Australian National University), “*Staffing the Digital Library in the 21st Century*,” at <http://anulib.anu.edu.au/>
- Steele, Colin (1997). *Managing Change in Digital Structures*, in Raitt, D. Libraries for the New Millennium (London, Library Association), pp148-168.  
[http://anulib.anu.edu.au/about/steele/managing\\_change.html](http://anulib.anu.edu.au/about/steele/managing_change.html)
- Websites:
- <http://jarda.cdlib.org/> An overview of the University of California's collection of historical records, personal artifacts, oral histories and photographs
- [http://www.jacar.go.jp/asia\\_en/index\\_en.html](http://www.jacar.go.jp/asia_en/index_en.html)
- <http://www.library.uiuc.edu/>

<http://www.usgenweb.org/>

Free genealogy and family history online made possible by the USGenWeb Project volunteers.

<http://www.fivecolleges.edu/> A consortium of colleges in the Pioneer Valley: Amherst, Hampshire, Mount Holyoke, and Smith Colleges, and the University of Massachusetts Amherst.

<http://www.oac.cdlib.org/> This union database of manuscripts and images from institutions across California is comprised of over 5000 archival finding aids. The OAC brings together historical materials from a variety of California institutions, including museums, historical societies, and archives. Over 120,000 images; 50,000 pages of documents, letters, and oral histories; and 8,000 guides to collections are available.

<http://www.library.cmu.edu/Research/ArchArch/Charette/> Online access to *Charette* magazine (vols. 1-54, 1920-1974) A project of the Carnegie Mellon University

<http://www.leeds.ac.uk/cedars/> Cedars began in April 1998 and ended in March 2002. Its broad objective was to explore digital preservation issues. These range through acquiring digital objects, their long-term retention, sufficient description, and eventual access.

<http://www.elib.gov.ph/>

<http://www.mainlib.upd.edu.ph/>

<http://www.supremecourt.gov.ph/elibrary/library.htm>

<http://www.ipophil.gov.ph/> (Intellectual Property Office official website)