Why preserve digital material?
the past, the present and the future

HICSA Conference
24 November 2008

Amelia Breytenbach and Ria Groenewald
Department of Library Services
University of Pretoria
Timeline

300 B.C. Library of Alexandria (Greece)
720 to 705 B.C. Library of Nineveh (Babylonia)
200 B.C. First Public Library in Rome (Italy)
1653 A.D. Library of Sobonne (France)
1973 A.D. British Library (England)
1993 A.D. Mosaic Web Browser 1.0 Invented
Emergence of Web 1.0
2004 A.D. Web 2.0 Blogging, Wiki, Social Networking
183 A.D. TCP/IP WAN Network Operational

Demetrius Phalereus (350 - 280 B.C.)
The inspirer of the foundation of the Ancient Library of Alexandria
The Municipal Building containing the Carnegie Library was constructed in 1903 and destroyed by fire in 1907

1992, May - August
• Sarajevo Library, a symbol of shattered culture
• Oriental Institute shelled and burned by Serb Nationalist forces

http://www/lisnews.org/node/28352

Photo 2007 – Library still not reconstructed
1998, July

Libraries under attack in Kosova

http://www.ifla.org/faife/faife/kosova/kosorepo.htm

The remains of the main public library in the municipality of Malishevë (South-Western Kosovo), which was burned down, July 1998.

The Prishtina University Library, 2008 (Flickr)
2001, September 11

Terrorist attacks in the USA


http://www.cathousechat.com/cathouse_chat/WindowsLiveWriter/TwinTowers911.jpg
www.arsenalofhypocrisy.com/.../image015.jpg
2002

Prague, Institute of Archaeology, Library

2002

Attack on The Polytechnical of Kabul's Library  (Photo: Artifa Rawan)

•http://excaliburrd.com/cs/blogs/excalibur/Securing%20Kabul.jpg
2003

Attack on Iraq’s Libraries

• http://excaliburrd.com/cs/blogs/excalibur/Securing%20Kabul.jpg

An assistant imam checks books taken to a mosque for safekeeping after reportedly being looted from Iraq's national library after the fall of Baghdad. Photograph: Saurabh Das/AP

What is left of Iraq's libraries
http://www.newpages.com/unclefrank/graphics/rumsfeld3opt.jpg
2004, September 09
Anna-Amalia Library, Weimar
central Germany

• http://palimpsest.stanford.edu/byform/mailing-list/book
arts/2004/09/msg00015.htm
2005, August

Hurricane Katrina

• http://www.regent.edu/general/library/about_the_library/news_publications/images/katrina1.png
• http://msnbcmedia2.msn.com/j/msnbc/Components/Photos/060622/060622_library02_hmed_6p.h2.jpg
2006, December 26

Tsunami (Sri Lanka)

- CDNLO Newsletter:

Inside view of the Ahangama public library after the tidal wave

The entire library building washed away by tsunami in Baticaloa, Sri Lanka

The ruined newspaper collection of the Magalle branch of the Galle public library

After the destruction, Hikkaduwa Public Library
2007, November 27
Street riots north of Paris destroyed the Louis Jouvet Library

http://www/lsnews.org/node/28352

Photos
• img.dailymail.co.uk/i/pix/2007/11_04/parisrio..
Street riots in Paris
• The NY Times reports, "The Louis Jouvet library was torched."
Photo taken by Charles Platiau/Reuters
2008, July 24
Swan's Island Library in Maine burns to the ground

Photo
The digital world and the library

- The physical digitization into searchable digital text
- Online interactive library
- The role of the library as a preserver and storehouse of digital media
- The searchability of large amounts of digital data contextually, beyond keywords
Digital preservation

- Digital preservation
  - is a term used for the storage and
  - ongoing protection of digitally born or digitally created material

- Safeguarding digital material
  - for future usage
  - continued accessibility
  - maintenance in the form of refreshing for future migration
  - enabling it to be accessed and reviewed by interested parties
    over a wide range of space and time
Why preserve?

- Fragile digital documents
- The records of the entire present period of history is in jeopardy
- Retaining digital information requires constant human attention
- You don’t know you have a gap in your digital records until you try to review something
Card catalogue

1. The card catalogue with cross referencing
2. Strong reliance on Dewey system
3. MARC21 records
4. AACR2/RDA
Definition of metadata

Conventional metadata
- Is data about a digital object
- Explain the technical method of creation and administrative data
- Contain the descriptive information

Preservation metadata
- Include data to authenticate the provenance of a digital object; and
- Contain a complete dataset to ensure the future usage of a digital object
Preservation metadata categories

Preservation Description Information (PDI)
- Reference
- Context
- Provenance
- Fixity
Description:

Colour photos.

Photo 1: Original document size: (w)7 x (h)4.64 cm. Original scanned size: 274 kb JPEG, 600 dpi. Final web-ready size: 26.29 kb. Estimate download time: 10 sec. @ 28.8 kbps.

Photo 2: Original document size: (w)7 x (h)4.62 cm. Original scanned size: 221 kb JPEG, 600 dpi. Final web-ready size: 24.86 kb. Estimate download time: 10 sec. @ 28.8 kbps.

Original TIFF file housed at the Dept. Veterinary Tropical Diseases, University of Pretoria. Metadata assigned by Prof. R.C. Tustin, Professor Emeritus: DVTD. His academic and professional experience includes: veterinarian for 54 years, senior lecturer at UP for 7 years, head of Department at UP for 17 years and Veterinary Council for 3 years.

Description Provenance:

Original scanned in at 600dpi 100% TWAIN scanning program using an EPSON 1640XL scanner. Downsized to 400 pixels in width, resolution 300dpi. Web version done automatically by PhotoShop 7 Software. Downloading time shift between 4 to 7 seconds. Date done April 2005. Original scanned in at 600dpi 100% TWAIN scanning program using an EPSON 1640XL scanner. Downsized to 400 pixels in width, resolution 300dpi. Web version done automatically by PhotoShop 7 Software. Downloading time shift between 4 to 7 seconds. Date done April 2005.
Need for Preservation Metadata

- Identifies the record
- Determines who created it
- Details the content of the digital object
- Puts the record in a context
- Provides technical details
- Provides knowledge about the bitstream and how it was created

Submitted by Amelia Breytenbach (abreyten@op.up.ac.za) on 2007-10-11T07:47:42Z
No. of bitstreams: 6
AT_fotoalbum6.jpg: 18703 bytes, checksum: b34593b272cb8fc200ade463859ed82d (MD5)
AT_fotoalbum5.jpg: 17862 bytes, checksum: ad287e77f2a71ffaa8eeb45acf7bfd27 (MD5)
AT_fotoalbum4.jpg: 17339 bytes, checksum: 4ce10ea57074f0b69daf1ad64923ac72 (MD5)
AT_fotoalbum3.jpg: 21651 bytes, checksum: 9b2d12d4355caa0a60033bd4e2d2b9a9 (MD5)
AT_fotoalbum2.jpg: 14891 bytes, checksum: ab557dff30175fde5f6ccf35b1ea690 (MD5)
AT_fotoalbum1.jpg: 12779 bytes, checksum: fefefe9f90df44fd483d7fe8322755f7 (MD5)
Standards

- **OAIS (Model widely adopted as starting point in digital preservation efforts)**
  (http://en.wikipedia.org/wiki/Open_Archival_Information_System)

- **PREMIS (Preservation metadata)**
  (http://www.oclc.org/research/projects/pmwg/premis-final.pdf)

- **Z39.87 (Technical metadata)**

http://www.lockss.org/lockss/OAIS
PREMIS MODEL

Intellectual entity (photo)

权利 = 对象 - 指令用户代表什么

转化成数字对象

TIFF图像文件

权利 = 对象 - 指令用户代表什么

转换为JPEG用于Web显示

Agent:
- 事件执行人的角色（姓名/组织）
- 软件名称和版本号
- 操作系统类型

Object:
- 文件大小
- 创建日期
- 文件格式
- 创建应用程序

Rights:
- 许可协议
- 授予的特定权限

保存以实现互操作性、访问和可读性
Creators of preservation metadata

- Creators of digital resources
- Digitization projects
- Fixity information

Object becomes useful and can be preserved for future usage

Descriptive and Preservation Metadata
The DCC Curation Lifecycle Model

www.dcc.ac.uk
info@dcc.ac.uk
Although several collections have been digitized and made available in the University of Pretoria’s Institutional Repository, a pilot study has not been done to measure the project management and workflow of future projects. There was a need to identify a project small enough to conform to normal project management requirements to use as an example to establish the planning and workflow of future projects. This paper offers practical help to libraries starting with digitization, it supplies valuable information for project management, planning of workflow and estimate time frames for completing a specific task in the digitization process.

**Title:** The African Elephant: a digital collection of anatomical sketches as part of the University of Pretoria’s Institutional Repository - a case study

**Creator:** Groenewald, Ria

Breytenbach, Amelia

**Description:** Although several collections have been digitized and made available in the University of Pretoria’s Institutional Repository, a pilot study has not been done to measure the project management and workflow. The collections available in the repository at the time of this project were all long-term projects. There was a need to identify a project small enough to conform to normal project management requirements to use as an example to establish the planning and workflow of future projects. This paper offers practical help to libraries starting with digitization, it supplies valuable information for project management, planning of workflow and estimate time frames for completing a specific task in the digitization process.

**Publisher:** The Authors

**Date:** 2007-11-30

**Type:** Article

**Format.extent:** 3.62 MB (3,796,480 bytes)

**Format:** MS Word 2003

**Resource ID:** 2007_gro_bre

**Language:** English

**Rights:** University of Pretoria

**Keywords:** Digital storage ; Collections management ; University libraries ; Anatomical drawings ; South Africa

**Abstract**

**Purpose:** Although several collections have been digitized and made available in the University of Pretoria’s Institutional Repository, a pilot study has not been done to measure the project management and workflow at the time of this project. There was a need to identify a project small enough to conform to normal project management requirements to use as an example to establish the planning and workflow of future projects. This paper offers practical help to libraries starting with digitization, it supplies valuable information for project management, planning of workflow and estimate time frames for completing a specific task in the digitization process.

**Design/methodology/approach:** A collection of anatomical sketches in the field of Veterinary Science, Department of Anatomy and Physiology was identified as an ideal project to conform to normal project management requirements. The collection consists of sketches made by the veterinary students participating in the program. The sketches are divided into categories such as musculature, bone, and skin. The sketches are digitized and then made available in the repository.

**Findings:** The completed project delivered on key areas such as the electronic and physical aspects through metadata creation. Basic preservation of the physical collection was ensured as well as the digital collections were archived for future use. The lessons learned and how it can be applied in future projects to the advantage of the library and the institution.

**Practical implications:** The paper provides a very useful case study for other libraries to develop their own digital collections.

**Originality/value:** This paper offers practical help to libraries starting with digitization, it supplies valuable information for project management, planning of workflow and estimate time frames for completing a specific task in the digitization process.

**Article Type:** Case study
Chaos of the ‘structured web’
Risk analysis for digital objects

- Hard drive failure
- URL error – linked broken
- Storage medium failure
- Loss of information/data
- Human error and memory
- Hackers
A preservation strategy is needed to safeguard digital content for future access.

A true preservation strategy must put planned-out business rules behind storage migration.

It should address:
- the risk factors involved
- the actions needed for digital preservation
- when is the estimated time to do so
- responsible entities [who are involved (IT/Librarians/Preservator)]
Actions required for digital preservation

- Store multiple copies
- Characterize and validate
- Allocating unique persistent identifiers with comprehensive metadata
- Develop and execute preservation plans
- Implement a comprehensive technology watch mechanism
- Develop or acquire tools for preservation actions
Successful digital preservation

- Longevity
- Interoperability
- Total cost of ownership
- Technology obsolescence protection
- Back-up and recovery support
Technical developments

- The technical challenges, like the perceptual issues, will be overcome with time.
- Bandwidth will increase.
- Processors will get faster.
- Memory will continue to grow.
- The two major technical hurdles that need to be addressed are:
  - the tools that are available for digital preservation.
  - interoperability.
“Change can be perceived, but not experienced. It is difficult, usually impossible to be aware of change while it is happening – but it is easy to see it has happened when one looks back”

T S Elliot
Starry, starry night.
Paint your palette blue and grey,
Look out on a summer's day,
With eyes that know the darkness in my soul.
Shadows on the hills,
Sketch the trees and the daffodils,
Catch the breeze and the winter chills,
In colors on the snowy linen land.
Starry, starry night.

Now I think I know what you tried to say to me,
How you suffered for your sanity,
How you tried to set them free.
They would not listen, they're not listening still.
Perhaps they never will...
Amelia Breytenbach
Metadata Specialist
Veterinary Library
University of Pretoria
Email: amelia.breytenbach@up.ac.za
Tel: 012 x 529-8391

Ria Groenewald
Digitization Coordinator
Merensky Library
University of Pretoria
Email: ria.groenewald@up.ac.za
Tel: 012 x 420-3792
References

- Katrina
  http://waynecountyreadingcouncil.com/yahoo_site_admin/assets/images/katrina_library.236181047_std.jpg
- Carnegie Library destroyed by fire, ...
  The building so recently constructed in 1903 was destroyed by fire in 1907.
- Sarajevo: On August 27, 1992. newcombat.net/article_thelibraries.html
  http://digitalpreswhitepaper.pdf
- An Introduction to Digital Preservation (TASI: http://www.tasi.ac.uk/advice/delivering/digpres.html)
- PREMIS (http://www.oclc.org/research/projects/pmwg/premis-final.pdf)