The Future is in Preservation Metadata

TLA Annual Conference April 1–4, 2003 Houston, Texas

Daniel Gelaw Alemneh
University of North Texas
Digital Preservation

Digital Preservation refers to series of managed activities necessary to ensure continued access to digital materials for as long as necessary.

“...the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable”

Margaret Hedstrom, (University of Michigan)
Why Digital Preservation Planning Now

♣ Proliferation of Digital Resources (DR)
  – 2 exabytes (10^9 bytes) of unique information per year.

♣ More and more DR are being created
  – Born digital resources
  – Various digitization initiatives

♣ User expectations
  – Anything, anytime, anywhere
“Digital Resources last forever – or five years, whichever comes first” (Rothenberg 1998)

- Most libraries have DR created 5 or more years ago in forms that can’t be accessed today

DR need to be managed as early as possible

- From the point of creation stage

If we wait longer to begin, our task increases in difficulty
Life Cycle of Digital Resources

**Creation**
- Digitizing analogue originals
- Born Digital

**Organization**
- Dissemination

**Preserve**
Life Cycle Management

- Creation
- Organization
- Dissemination
- Preserve
Issues in Digital Preservation

♣ Technological issues

- Inadequate media longevity
- Rapid Technology (hardware, software) obsolescence
- Heterogeneity

♣ Organizational issues

- Cost
- Expertise
- New Roles & organizational structure

♣ Legal issues

- Copyright & IPR
- Licensing
- Privacy and Confidentiality

♣ Other issues

- Access vs. Security
- Authenticity
Current Preservation Strategies

♣ Creating multiple copies in different media
  o Geographically separate storage

♣ Refreshing
  o Copying DR from one to another storage medium of the same type

♣ Technology preservation
  o Technology museum

♣ Migration
  o The periodic transfer of DR from one hardware/software configuration to another or from one generation of technology to a subsequent one

♣ Emulation
  o Programs that mimic the behavior of the original technical environment (See CAMiLEON Project)
Preservation Strategies: Migration Over Time

Time-1
User-1
Texas Register Original

Time-2
User-2
Texas Register Migrated

Migration Tool - i

Time-3...n
User-3...n
Texas Register Migrated

Migration Tool - 2
Preservation Strategies:
Migration

Reversible Migration

Irreversible Migration (loss Proven)
Responses at Various Levels

♣ The current preservation strategies are far from perfect
♣ Effective metadata has an important part to play in any successful preservation strategy
♣ Various National and International Initiatives
Preservation Metadata Projects

- **NEDLIB**: (Networked European Deposit Library)  
  [http://www.kb.nl/coop/nedlib/](http://www.kb.nl/coop/nedlib/)
- **NLA**: (National Library of Australia)  
- **PANDORA**: (Preserving and Accessing Networked Documentary Resources of Australia)  
- **British Library**:  
  [http://www.bl.uk/services/preservation/national.html](http://www.bl.uk/services/preservation/national.html)
- **Cedars** (CURL exemplars in digital archives)  
  [http://www.leeds.ac.uk/cedars](http://www.leeds.ac.uk/cedars)
- **RLG/OCLC**:  
  [http://www.oclc.org/research/pmwg/](http://www.oclc.org/research/pmwg/)
"Preservation Metadata is the information necessary to carry out, document and evaluate the processes that support the long-term retention and accessibility of digital content."

(OCLC/ORG)
UNT’s Approach

♣ The UNT digital preservation projects based on a realistic heterogeneous digital collections:

– Federal Documents Projects
  u CyberCemetery
  u WWII posters etc.

– State Documents Projects
  u Texas Register
  u Gammel’s Laws of Texas
  u Texas Electronic Depository Library Project
  u Portal to Texas History Project

– UNT Libraries Project
  u Government documents, Media library, Music Library, Rare Books, Archives, Dissertations and Theses, etc.
UNT’s Approach

▪ Considers many strategic and practical issues associated with digital preservation
  - Creating rich digital masters
  - Creating multiple backup copies
  - Formulating preservation policy
  - Documentation of processes
  - Building a technical structure to support preservation activities

♣ Acknowledge the centrality of metadata in the digital preservation process
  - Developing a specification for preservation metadata
    - Based on OAIS, Cedars, RLG/OCLC, NLA, and DC models
    - Draft preservation metadata elements at:
      http://texashistory.unt.edu/Draft_Metadata_Elements.html
Main parts of UNTL metadata:

- Resource discovery metadata
  - DC Based
- Technical metadata
  - Technical details about the original hardware and software environment
- Rights management metadata
  - Intellectual property rights, ownership, deposit agreements, etc.
- Preservation description metadata
  - Preserving integrity and authenticity
Preservation Metadata

PRESERVATION METADATA

DR Description Metadata  Technical Metadata  Administrative Metadata  Preservation Description

Software  Hardware  Right Mg't  IP  Metadata Mg't
Input Guidelines

♣ Each element is assigned:
  - **Label**
    - Name of the element
  - **Sub-elements**
    - Indicate existence of sub-elements
  - **Origins**
    - Source of the element
  - **Definition**
    - A brief statement that defines the element
  - **Description**
    - Further description for clarification the purpose of the element
  - **Examples**
    - Local application examples
  - **Obligation**
    - Indicate whether the element value is mandatory or optional
  - **Repeatable**
    - Indicate whether the element is repeatable or not repeatable
  - **Comments**
    - Notes to clarify exceptions
Example

♣ See the UNTL Sample Metadata Template

The PORTAL to TEXAS HISTORY
Embark on a Voyage of Discovery
http://texashistory.unt.edu
Example

```xml
<?xml version="1.0"?>
<!-- $Id: sample.xml,v 1.4 2003/03/29 15:34:35 Daniel Exp $ -->
<UNTL>

<!-- 1: title -->
<title>Eric the half a Bee</title>
<title role="alternate">The Fish Licence Song</title>
<title role="series">Monty Python's Flying Circus</title>

<!-- 2: collection -->
<collection>Humor</collection>

<!-- 3: creator -->
<creator>
  <name>Moen, Bill</name>
  <name>Moen, William E.</name>
  <information>wears a hat</information>
  <role>author</role>
</creator>
<creator>
  <name>Hartman, Cathy</name>
  <information>Head, Government Document Department-UNT</information>
  <role>editor</role>
</creator>
```
Example

- <!-- 4: subject -->
  - <subject authority="LCSH">comedy songs</subject>
  - <subject authority="Sears">humorous songs</subject>
- <!-- 5: description -->
  - <description>
    - This is the song that John Cleese sings to Michael Palin after
      the latter has refused to sell a fish licence to the former.
  - </description>
- <!-- 6: date -->
  - <dateCreated>1972</dateCreated>
  - <dateDigitized>1998-03-12</dateDigitized>
- <!-- 7: publisher -->
  - <publisher>Leonard Corp.</publisher>
- <!-- 8: contributor -->
  - <contributor role="singer">John Cleese</contributor>
  - <contributor role="listener">Michael Palin</contributor>
  - <contributor role="chorus">The Fred Tomlinson singers</contributor>
- <!-- 9: source -->
  - <source>Monthly Python's Contractual Obligation Album</source>
- <!-- 10: identifier -->
  - <identifier type="url">http://www.python.net/script/eric.html</identifier>
  - <identifier type="ISBN">123-456-789-01</identifier>
  - <identifier type="callNumber" subtype="LC">M1507.B47</identifier>
  - etc.
- </UNTL>
Next Step

♣ Completion of an XML environment for management and production of UNTL metadata
♣ Finalize standards and best practices that offer guidance with preservation and metadata management.
♣ Identifying responsibility for each activity
♣ Mapping to MARC and other widely used standards
Concluding Thoughts

♣ There is no access without preservation.

♣ Digital preservation must become part of the core mission of libraries. It needs to begin now and continue, and the strategies need to be a combined and flexible approaches to accommodate unforeseen factors.

♣ Metadata provides a critical part of the solution to the preservation challenges including detecting preservation threats and promoting preventive measures.

♣ Successful “preservation strategy” requires Commitment to working collaboratively (regionally & internationally).

♣ The promise of long term access is at hand. In deed, the future is in preservation metadata!!!
Thank You!

Questions?

Dalemneh@library.unt.edu