

The Cybercemetery: Prolonging Usable Afterlife

Cathy Nelson Hartman

Digital Projects Department, University of North Texas, Denton, Texas

Samantha Kelly Hastings

School of Library & Information Sciences, University of North Texas, Denton, Texas

Daniel Gelaw Alemneh

Doctoral Student in Information Science, University of North Texas, Denton, Texas

Abstract

This paper discusses issues related to digital resources management when capturing and preserving Web-based, heterogeneous digital materials produced by a variety of software in various versions. Despite the current inadequate digital preservation solutions, the writers explore various methods and tools that facilitate the efficient management of vast quantities of dynamic and heterogeneous digital information resources. The *CyberCemetery* project at the University of North Texas is used as an example of this type of endeavor. It specifically demonstrates the efforts being made by UNT to implement digital preservation strategies for prolonging the usable life of such heterogeneous digital resources. This paper also highlights the potential role of metadata at all levels in the life cycle of a digital resource (creation, management, use, and preservation).

Introduction

The University of North Texas (UNT) is one of over 60 federal depository libraries in Texas and 1,250 in the nation, storing and providing free public access to government information. In the 1990s, many government agencies and commissions changed their publishing practices and began providing access to their publications in a digital format over the World Wide Web. Since 1997, the UNT Libraries has formed partnerships with state and federal agencies to preserve their electronic publications.

The CyberCemetery Project

One such partnership is the *CyberCemetery*. In an agreement with the U.S. Government Printing Office (GPO), UNT preserves the Web sites of federal agencies

that have ceased operation. This project presents many challenges related to the capture and long-term preservation of this heterogeneous collection.

In 1997, a Memorandum of Understanding was signed by UNT and the U.S. Government Printing Office (GPO) to outline respective roles in the preservation of the electronic publications of a government agency that had ceased operation, the Advisory Commission on Intergovernmental Relations. This was the first “dead” agency ingested into the site, which soon became known by its users as the “CyberCemetery.” It presently provides public access to the electronic Websites of 22 “deceased” agencies and preserves the materials for future access. At this time, work is underway to ingest 15+ government agencies that were folded into the complex structure of the new Homeland Security Department.

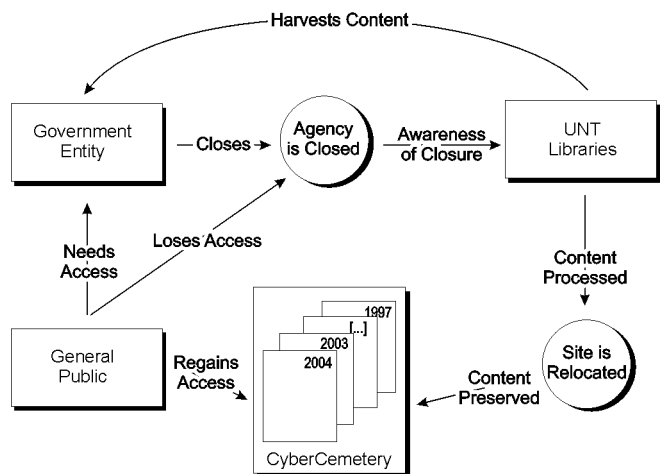


Figure 1. Framework for Cybercemetery Project's processes and relationships

According to the user statistics, most of the sites receive as much or more traffic now than they did when the agency was alive. Popular defunct agency sites available at the *Cybercemetery* include: the National Partnership for Reinventing Government (the then-Vice president Al Gore initiative, which expired in early 2001 with the change in administrations) and the Office of Technology Assessment, (which closed September 29, 1995, after 23 years of analyzing the public policy implications of scientific and technological innovations). For complete descriptions of the agency sites included, please visit the *CyberCemetery* at: <http://govinfo.library.unt.edu/>



Figure 2. *Cybercemetery* Project Page

Challenges

In recent years, our entire information environment has changed and continues to change at a breathtaking pace. Digital technologies are shaping creation, management, access, and preservation in ways, which are so profound that traditional methods no longer are effective. Increasingly complex Web sites, for instance, increased the difficulty of the capture, ingest, and preservation process. Additionally, file types ranged from plain text to multimedia materials, and files were created in a variety of proprietary software.

The various technology implementations in the Web environment increases the level of complexity for capture and preservation. As the volume and complexity of resources increased, the need for highly-developed resource management tools became apparent. Several efforts have been made by various communities to integrate the various information sources scattered across several environments by providing interoperability between the multiple sources and systems. In recent years, many of the information management system vendors are migrating toward a client-server architecture. Such an architecture facilitates distributed information management capabilities by enabling users of one system to access other systems in an efficient manner.

The issues of long term access and preservation represents a complex set of challenges, which are exceptionally difficult for institutions to address individually. New architectures and tools must be created to enable peer-to-peer collaboration on such important issues as geographically separate back-up storage, metadata creation and sharing, and implementation of preservation strategies.

Long-term preservation planning for such diversified digital resources, which were created in a variety of hardware/software environments, must include policy decisions regarding the level of functionality that should and can be preserved. Among other considerations, responsible and viable preservation activities need to address issues such as:

- Is access to the full range of functionality and content important, or is simple access the 'bare bones' intellectual content enough?
- Can we (technologically, economically, legally, etc.) ensure the preservation of digital resources at a level acceptable to users?

The current library and information science literature reflects the preservation issues as one of the most pressing needs and challenges of digital resource management. Several of the most influential and prolific digital library researchers, such as Hedstrom (2002) and Bessar (2002), agree that the current digital preservation solutions are inadequate to accommodate the complex set of challenges brought by dynamic, heterogeneous digital resources. However, there is a growing body of research emphasizing the critical role of metadata as an enabling tool in any successful preservation strategy. Like many others, UNT advocates a metadata-based approach to the preservation of digital resources.

Accordingly, UNT is building an infrastructure to assist with preservation for long-term access to its digital resources. An application framework, based on open source software, was created, and implementation of a metadata based preservation strategy is underway. A preservation metadata element set was derived by merging various metadata standards. A detail description of the UNT metadata element set may be viewed at: <http://texashistory.unt.edu/guides/help/UNTL-Metadata-Guide.htm>.

This metadata logs the strategic context for the creation, management, preservation, and use of digital collections. Adequate information about the life cycle and technical environment needed to view the digital resources supports our preservation activities including the widely implemented strategies of "migration" (transfer of digital resources from one generation to a subsequent generation) and "emulation" (developing techniques for imitating obsolete systems on future generations of computer), among others.

Conclusion

In today's digital libraries environment in which diverse collections are integrated and linked from a single access point, the longevity issue is one of the key challenges facing any library and information system. Like many others, the *CyberCemetery* project staff is working to create needed tools to address the preservation and long-term access issues for its ever expanding digital collections. Recognizing the fact that any long-term solution is likely to depend on collaborative arrangements, UNT is forging partnerships with other institutions for tool development and to establish coordinated action. Digital preservation activities and methods must be automated and shared among peer institutions to reduce costs if the vast quantities of digital information are to be preserved for long-term access.

References

1. Wallace Koehler, A longitudinal study of Web pages continued: a report after six years. *Information Research*, 9 (2) paper 174 (2004). [Available at: <http://InformationR.net/ir/9-2/paper174.html>].
2. Daniel Alemneh, Samantha Hastings, & Cathy Hartman, A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience. *FirstMonday*, 7(8). (2002). [Available at: http://www.firstmonday.dk/issues/issue7_8/alemneh/index.html].
3. Howard Besser, The next Stage: Moving from Isolated Digital Collections to Interoperable Digital Libraries. *FirstMonday* 7(6). (2002). [Available at: http://www.firstmonday.dk/issues/issue7_6/besser/index.html].
4. Margaret Hedstrom, Research Challenges in Digital Archiving: Towards a National Infrastructure for long-term Preservation of Digital Information (2002). [Available at: <http://www.si.umich.edu/digarch/>].

Biography

Cathy Nelson Hartman, initiator of the CyberCemetery Project, is Head of the Digital Project Department at the University of North Texas and Associate Fellow of the Texas Center for Digital Knowledge. Hartman actively participates at the state and national level, most recently chairing an ALA committee on digitization of government information and chairing the Depository Library Council, a federal government advisory board. She was recently elected to the Texas Records Management Interagency Coordinating Council. She publishes and speaks widely and is a successful grant recipient.

Dr. Samantha .Kelly Hastings joined the faculty at the UNT in 1995. She is very active in state and national professional associations and she is the current president of ASIS&T. Dr. Hastings has served as a resource person and presented a number of papers at various professional meetings and conference programs. She was a principal investigator for a federally funded IMLS Library- Museum-University Collaboration project. Dr. Hastings continues to research problems associated with the access, retrieval, and preservation of digital images, with particular emphasis on designing information communities for the 3D environment.

Daniel Gelaw Alemneh is a doctoral student in Information Science at the University of North Texas. He received his Post Masters in Digital Image Management from the University of North Texas in 2000. Daniel earned his Masters in Information Management from the University of Sheffield, UK in 1997. Mr. Alemneh is employed as a Senior Graduate Library Assistant in the Digital Project Department of the UNT, and works on various digital projects.

Abstract Title—Sample Page

Cathy Nelson Hartman

Digital Projects Department,
University of North Texas
Libraries
Denton, Texas

Samantha Kelly Hastings

School of Library &
Information Sciences
University of North Texas,
Denton, Texas

Daniel Gelaw Alemneh

Doctoral Student,
Interdisciplinary Program in
Information Sciences,
University of North Texas
Denton, Texas

Abstract

This paper discusses issues related to digital resources management when capturing and preserving Web-based, heterogeneous digital materials produced by a variety of software in various versions. Despite the current inadequate digital preservation solutions, the writers explore various methods and tools that facilitate the efficient management of vast quantities of dynamic and heterogeneous digital information resources. The *CyberCemetery* project at the University of North Texas is used as an example of this type of endeavor. It specifically demonstrates the efforts being made by UNT to implement digital preservation strategies for prolonging the usable life of such heterogeneous digital resources. This paper also highlights the potential role of metadata at all levels in the life cycle of a digital resource (creation, management, use, and preservation).

Keywords

Preservation, Longevity, Metadata, Long-term access, Digital resource management,