



Collection Plan
for
the CyberCemetery

8.21.2006

Prepared by:

Valerie Glenn
Starr Hoffman

Contents

Section 1.	Mission & Scope	3
Section 2.	Selection	5
Section 3.	Acquisition	6
Section 4.	Descriptive Metadata.....	7
Section 5.	Presentation and Access	8
Section 6.	Maintenance and Weeding	10
Section 7.	Preservation.....	11
Appendix A.	Memorandum of Understanding	12
Appendix B.	Web Archiving Service Agreement	13
Appendix C.	Seed List	14
Appendix D.	UNT Metadata Schema (Descriptive Metadata)	17
Appendix E.	UNT Metadata Schema (Preservation Metadata)	19

Section 1. Mission & Scope

A. Mission Statement

The University of North Texas Libraries Mission

The mission of the University of North Texas Libraries is to acquire, preserve, provide access to, and disseminate recorded knowledge in all its forms. Access will be provided increasingly through electronic networks and consortial arrangements. The Libraries, through traditional methods and through digital information resources, provide bibliographic, reference, and instructional support to assist the university's programs of teaching, research, scholarly and creative production, and public service.

Mission as a Federal Depository Library

Federal depository libraries must maintain information products disseminated by the Federal Depository Library Program as part of their existing collections and are responsible for providing permanent public access to this material, including digitally scanned and "born digital" documents.

B. User Groups

Because the CyberCemetery is publicly accessible online, it serves a broad user group that encompasses the students, faculty, and staff of the University of North Texas, but also includes researchers, government employees, and librarians. Because the user group is so broad and the CyberCemetery covers so many different subjects, user information needs are many and varied. Two common information needs are examples of past policies and information about specific topics.

C. Collection Subject, Theme, or Event

The CyberCemetery is a collection of government websites that have ceased operation (usually websites of defunct government agencies and commissions that have issued a final report). This collection features a variety of topics indicative of the broad nature of government information. In particular, this collection features websites that cover topics supporting the university's curriculum and particular program strengths.

This collection supports the mission of the UNT Libraries by acquiring, preserving, and providing access to recorded knowledge in the form of government websites.

This collection supports the mission of the Federal Depository Library Program by providing permanent public access to government information in the form of websites.

D. Curator(s)

Starr Hoffman
Librarian for Digital Collections
University of North Texas Libraries
Government Documents
P.O. Box 305190

Denton, TX 76203-5190

SHoffman@library.unt.edu

940.565.4150

Fax: 940.565.2599

Suzanne Ebanues

Planning and Development Specialist

Office of Planning and Development, Library Services

U.S. Government Printing Office

732 North Capitol St., NW

Mail Stop: IDLP

Washington, DC 20401

sebanues@gpo.gov

202.512.2010, x-33269

Fax: 202.512.2300

Section 2. Selection

A. Seed List

Suitable websites must meet the following criteria:

- created by a government agency or commission
- be at-risk (i.e., suspected of being eminently removed or changed)
 - defunct (or soon-to-be) government agency
 - commission has completed final report
 - website with material that frequently changes or has been/will be edited (previous information not saved)
 - website which has been untended for some time and suspected to be no longer active (may be decommissioned at any time)
- support the UNT curriculum and/or the needs of the community
 - example: topics relevant to UNT programs, such as geology, history, criminal justice, public administration, etc.
 - example: topics relevant to community members, such as tax reform, bankruptcy, etc.

The materials included in this collection are listed in Appendix C: Seed List (8.7.06 report). In the future, this list should be generated from the Web Archiving Service (WAS) and should be updated as new websites are captured.

B. Capture Scope

In general, the capture scope for websites in the CyberCemetery will be the entire site (i.e., materials or content originating from the host specified in the seed URL). However, some sites may also necessitate capturing one level (link-hop) outside of the domain. Most of the websites will be hosted on .gov domains; however, there are exceptions to this rule (e.g., some government websites or specific media on some of these sites may be hosted by commercial providers). Because of these and other exceptions, each URL on the seed list must be assessed individually.

C. Rights Management

In general, websites in this collection are copyright-exempt because they are produced by the federal government.

All materials in this collection are covered by the legal agreement between the Government Printing Office and the University of North Texas. For more information on this agreement, see Appendix A: Memorandum of Understanding.

Section 3. Acquisition

A. Frequency of Capture

Websites for this collection will be captured once. They will be captured when a government agency becomes inactive, when an expiration date has passed, and/or after a final report is issued. They will be recaptured only if information is determined to be incomplete.

Note: It is important to note that sometimes a site will change while it is being harvested. Pages may be removed, moved, or may be changed by the content provider during the capture. This could result in hyperlink errors or semantic inconsistencies among pages of a captured web site when subsequently viewed by users.

B. Capture Scope

[This section is not relevant to the collection.]

C. Material Types & Formats

There are no specific material types or formats to be excluded from this collection. In general, complete websites are to be captured, encompassing files of all types included on their domain. However, if a website is particularly large, the curator should discuss its capture with ITS (Information Technology Services) before capturing it.

D. Interactive & Dynamic Content

The legal agreement regarding this collection specifies that captured websites must be represented as exactly as possible within the archive. Exceptions that necessitate some minor changes to the HTML are:

- Add this text to comment forms/contact pages: "Please do not send comments. [Enter date of final commission report, if applicable.] We archived this site in [date]. If you experience any technical problems using this site, please let us know [UNT email link here]." (If it is a website for a commission, insert date that final report was submitted.)
- Email links: preserve the display of email addresses, but disable links
- Submit buttons: preserve their display, but disable
- Also, add the word "Archive" (in 8 pt. Times New Roman font and center-justified) to every page of each archived website.

Section 4. Descriptive Metadata

A. Level of description

Website level description provides ample subject access to the information.

B. Metadata elements

Our local metadata schema is listed in Appendix D: UNT Metadata Schema (Descriptive Metadata).

C. Controlled vocabularies

The CyberCemetery collection will use the Legislative Indexing Vocabulary (LIV) for subject terms. The Legislative Indexing Vocabulary is a thesaurus of over 5,000 terms created by the Congressional Research Service to describe legislative and public policy literature.

Section 5. Presentation and Access

A. Discovery

The web interface to the CyberCemetery collection is provided by the University of North Texas Libraries' Government Documents Department <<http://govinfo.library.unt.edu/>>. The primary method of discovery is the search screen, which provides full-text searching. That search can be narrowed to search a few specific websites, such as the National Commission on Terrorist Attacks Upon the United States. The descriptive metadata includes subject terms that will be used to provide topical access (see Section 4). Users also have the option of browsing the collection by website name, branch of government, or date of expiration.

Users are likely to use both material titles and subject(s) to evaluate the results returned from a search of the collection.

B. Access

This web collection is visible. There are no time-dependent release restrictions associated with the materials in this web collection. We try not to restrict the accessibility of captured web content; however, there may be exceptions. In some cases, a physical residence address appearing on an archived website may be removed at the resident's request (this is done to protect the privacy of government employees and officials, as well as citizens who have publicly commented on agency proposals). In the case of such exceptions, explanatory text will be included on the page to explain the reason for the removal.

C. Look-and-Feel

It is important to the users of this collection to retain the "look-and-feel" of the source websites. Out of necessity for identification as archived websites, clarifying text notices will be inserted on comment pages and/or content forms, and a text notice reading "archive" will be inserted at the top of each homepage.

D. Dynamic Content

It is desirable that users be allowed to access hyperlinked materials and websites that are not located within the web archive. Users should be alerted to the fact that they are leaving the archive by an informative message, and should be provided both a link to the external website and a link back to the archived page.

Email links and form submit buttons should be displayed, but disabled. Informational text should inform the user that this website has been archived, and give contact information for the archiving institution, in case of technical difficulties with the site.

E. Multiple Types/Formats

All types and formats of materials in the captured content of this collection will be discoverable and accessible by all users regardless of their location, assuming that the users have access to the required software. The content will always be accessible at the University of North Texas Library, which will retain the required software for discovery and access.

F. Authenticity

Authenticity will be established by the CyberCemetery interface through which users will access the websites, and by the <http://govinfo.library.unt.edu/> URL stem which all these sites will share. The reputation of the University of North Texas and its legal agreement with the Government Printing Office and the National Archives and Records Administration will be a sufficient measure of material authenticity for the collection's users. This collection will not use digital stamps or watermarks as indicators of authenticity, because our legal agreement requires that the code be altered as little as possible, in order to best preserve the original "look-and-feel" of the websites. The only changes that will be made to the code are those which are necessary to indicate that the website is an archived page and those which disable actions such as forms and email links (for more on this, see Section 3, part D).

Section 6. Maintenance and Weeding

A. Maintenance Activities

Periodically, the curator will identify new websites to be added to the collection.

B. Deselection Guidelines

There are no anticipated situations under which deselection will be required; deselection is expressly denied in the legal agreement between the University of North Texas, the Government Printing Office, and the National Archives and Records Administration.

Some information may be removed by specific request, but entire websites will not be removed (see Section 5-B).

C. Collection Evaluation

[This section is not applicable to this collection.]

Section 7. Preservation

A. Technology Obsolescence

Preserving original hardware and/or software may not effectively preserve the collection, because the majority of users access the collection off-campus, using their own hardware and software. Emulation may be a better method; we are currently considering our options. For more on the terms of the preservation service, see Appendix B: Web Archiving Service Agreement.

B. Preservation Metadata

See Appendix E: UNT Metadata Schema (Preservation Metadata).

Appendix A. Memorandum of Understanding

[MOU text here]

Appendix B. Web Archiving Service Agreement

[text here]

Appendix C. Seed List

Seed List as of 8.7.2006

Access America (2001)

<http://govinfo.library.unt.edu/accessamerica/index.html>

Advisory Commission on Electronic Commerce (2000)

<http://govinfo.library.unt.edu/ecommerce/index.htm>

Advisory Commission on Intergovernmental Relations (1996)

<http://www.library.unt.edu/gpo/acir/acir.htm>

Advisory Commission on Online Access and Security (2000)

<http://govinfo.library.unt.edu/acoas/index.htm>

Advisory Committee on Public Interest Obligations of Digital Television Broadcasters (1998)

<http://govinfo.library.unt.edu/piac/pubint.htm>

Amtrak Reform Council (2002)

<http://govinfo.library.unt.edu/arc/>

Columbia Accident Investigation Board (2003)

<http://govinfo.library.unt.edu/caib/default.html>

Census Monitoring Board–Congressional Members (2001)

<http://govinfo.library.unt.edu/cmb/cmbc/index.htm>

Census Monitoring Board–Presidential Members (2000)

<http://govinfo.library.unt.edu/cmb/cmbp/index.htm>

Coalition Provisional Authority (2004)

<http://govinfo.library.unt.edu/cpa-iraq/index.html>

Commission on Affordable Housing and Health Facility Needs for Seniors in the 21st Century (2002)

<http://govinfo.library.unt.edu/seniorscommission/>

Commission on Structural Alternatives for the Federal Court of Appeals (1999)

http://www.library.unt.edu/gpo/csafca/app_comm_uscourts_gov.html

Commission on Review of Overseas Military Facility Structure of the United States: Overseas Basing Commission (2005)

<http://govinfo.library.unt.edu/osbc/>

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction (2005)

<http://govinfo.library.unt.edu/wmd/>

Commission to Strengthen Social Security (2001)

<http://govinfo.library.unt.edu/csss/index.htm>

- COPA (Child Online Protection Act) Commission (2000)
<http://govinfo.library.unt.edu/copacommission/>
- International Competition Policy Advisory Committee (ICPAC) Research Collection (2000)
<http://www.library.unt.edu/gpo/icpac/icpac.htm>
- Millennial Housing Commission (2002)
<http://govinfo.library.unt.edu/mhc/>
- National Bankruptcy Review Commission (1997)
<http://govinfo.library.unt.edu/nbr/index.html>
- National Bioethics Advisory Commission (2001)
<http://govinfo.library.unt.edu/nbac/nbac.html>
- National Civil Aviation Review Commission (1997)
<http://www.library.unt.edu/gpo/NCARC/index.htm>
- National Commission on Terrorist Attacks Upon the United States (2004)
<http://govinfo.library.unt.edu/911/>
- National Commission to Ensure Consumer Information and Choice in the Airline Industry (2002)
<http://govinfo.library.unt.edu/ncecic/>
- National Drought Policy Commission (2000)
<http://govinfo.library.unt.edu/drought/default.htm>
- National Education Goals Panel (NEGP) (2002)
<http://govinfo.library.unt.edu/negp/index.htm>
- National Gambling Impact Study Commission (NGISC) (2001)
<http://govinfo.library.unt.edu/ngisc/index.html>
- National Partnership for Reinventing Government (2001)
<http://govinfo.library.unt.edu/npr/default.htm>
- Office of Consumer Affairs (1999)
<http://www.library.unt.edu/gpo/oca/default.htm>
- Office of Independent Counsel Donald C. Smaltz in re Secretary of Agriculture Alphonso Michael Espy (2001)
<http://govinfo.library.unt.edu/oic/>
- Office of Technology Applications–Highway TechNet (1999)
<http://www.library.unt.edu/gpo/ota/index.html>
- Office of Technology Assessment (OTA) (1995)
<http://govinfo.library.unt.edu/ota/default.htm>
- President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry (1998)

<http://govinfo.library.unt.edu/hcquality/>

President's Advisory Panel on Federal Tax Reform (2005)

<http://govinfo.library.unt.edu/taxreformpanel/>

President's Commission on Improving Economic Opportunity in Communities Dependent on Tobacco Production while Protecting Public Health (2001)

<http://govinfo.library.unt.edu/tobacco/index.htm>

President's Commission on Moon, Mars and Beyond (2004)

<http://govinfo.library.unt.edu/moontomars/index.htm>

President's Commission on the Celebration of Women in American History (2001)

<http://govinfo.library.unt.edu/whc/whc.htm>

President's New Freedom Commission on Mental Health (2003)

<http://govinfo.library.unt.edu/mentalhealthcommission/index.htm>

Presidential Advisory Commission on Holocaust Assets in the United States (2001)

<http://govinfo.library.unt.edu/pcha/>

Smithsonian Institution Science Commission (2003)

<http://govinfo.library.unt.edu/sciencecommission/>

Special Oversight Board for Department of Defense Investigations of Gulf War Chemical and Biological Incidents (2000)

<http://govinfo.library.unt.edu/oversight/>

United States Commission on Ocean Policy (2004)

<http://govinfo.library.unt.edu/oceancommission/index.html>

United States Information Agency (USIA) [University of Illinois at Chicago] (1999)

<http://dosfan.lib.uic.edu/usia/index.html>

U.S. Commission on National Security/21st Century [Hart-Rudman Commission] (2001)

<http://govinfo.library.unt.edu/nssg/index.html>

U.S. Trade Deficit Review Commission (2000)

<http://govinfo.library.unt.edu/tdrc/index.html>

White House Commission on Complementary and Alternative Medicine Policy (2002)

<http://govinfo.library.unt.edu/whccamp/>

Appendix D. UNT Metadata Schema (Descriptive Metadata)

DESCRIPTIVE METADATA

Title

- Parallel Title
- Alternate Title
- Added Title
- Series Title
- Serial Title
- Uniform Title
- Other

Creator

- Creator Name
- Creator Role
- Creator Information

Contributor

- Contributor Name
- Contributor Role
- Contributor Information

Publisher

- Publisher Name
- Place of Publication
- Publisher Information

Date

- Original Resource Creation Date
- Digital Creation Date

Language

Description

- Content Description
- Physical Description

Subject and Keywords

Primary Source

Coverage

- Place Name
- Time Period
- Date*

Date Range

Source

Relation

Collection

Institution

Rights Management

Resource Type

Format

Identifier

URL

URN

DOI

ISBN

ISSN

OCLC No.

Report No.

Government Document No.

Accession or Local Control No.

UNT Catalog No.

RISM No.

Other Identifier

Note

Metadata Information

Metadata Creator

Date of Creation

Metadata Modifier

Date of Modification

Appendix E. UNT Metadata Schema (Preservation Metadata)

PRESERVATION METADATA
<i>File Information</i>
File Size
File Name
Format Name
Format Version
File description
Resolution
Dimension
Duration
Rate
Tonal-Resolution
Color
Compression
Other File information
Fixity Information
Authentication Type
Authentication Result
Date
<i>First Date</i>
<i>Last date</i>
System Information
Software
Creation Application Software
<i>Creation Application Name</i>
<i>Creation Application Version</i>
Access Application Software
<i>Access Application Name</i>
<i>Access Application Version</i>
Other Software Information
Hardware
Creation Hardware
Access Hardware

Other Hardware Information
Documentation
Structural Composition
Storage Medium
Access Inhibitors
Inhibitor Key
Functionality
Exception
Alteration History
Action Taken
Date of Alteration
Modifier
Other Alteration Information
Metadata Information
Metadata Editor/Modifier
Metadata Creation/Modification Date
Metadata Modification Action
Other Metadata Information
Comments