

# PRESERVING ACCESS TO GOVERNMENT WEBSITES

Development and Practice in the CyberCemetery

---

Starr Hoffman  
Librarian for Digital Collections  
University of North Texas Libraries  
13 August 2008

## PRESENTATION CONTENTS

---

- × Intro: *What is the CyberCemetery?*
- × Purpose: *Why create a CyberCemetery?*
- × Development
- × Archiving Process
- × Technical Details
- × Users by Country
- × Conclusion

# WHAT IS THE CYBERCEMETERY?


The screenshot shows the CyberCemetery website. At the top right, it says "University of North Texas Libraries". The main header is "CyberCemetery" with a speech bubble icon. Below the header is a search bar with a "Search" button. On the left side, there is a navigation menu with the following items: "CyberCemetery Home", "Search the CyberCemetery", "Browse the CyberCemetery", "Related Resources", "Contact Us", and "Digital Collections Home". Below the menu is contact information for the University of North Texas Libraries Government Documents: "P.O. Box 305190", "Denton, TX 76203-5190", "Phone: (940) 565-2870", and "Fax: (940) 565-2599". The main content area contains two paragraphs of text. The first paragraph states that the University of North Texas Libraries and the U.S. Government Printing Office created a partnership to provide permanent public access to the Web sites and publications of defunct U.S. government agencies and commissions. The second paragraph explains that the CyberCemetery is an archive of government websites that have ceased operation. At the bottom of the page, there is a footer with a navigation menu: "CyberCemetery Home | Search the CyberCemetery | Browse the CyberCemetery | Related Resources | Contact Us | Digital Collections Home | Government Information Connection | UNT Libraries | UNT". Below the footer, it says "This page is maintained by Starr Hoffman, and was last updated on November 08, 2007."

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

# WHAT IS THE CYBERCEMETERY?

- ✘ online archive of websites from U.S. government agencies or commissions that are no longer operating

This web site was frozen on September 20, 2004 at 12:00 AM, EDT. It is now a Federal record managed on behalf of the National Archives and Records Administration.  
External links were active as of that date and time. For technical issues, contact <https://www.library.unt.edu/forms/contact-us/cybercemetery/>.



### NATIONAL COMMISSION ON TERRORIST ATTACKS UPON THE UNITED STATES

About the Commission | Report | Hearings | Staff Statements | Press | Archive

The National Commission on Terrorist Attacks Upon the United States (as the 9-11 Commission), an independent, bipartisan commission created by congressional legislation and the signature of President George W. Bush in 2002, is chartered to prepare a full and complete account of the circumstances surrounding the September 11, 2001 terrorist attacks, including preparation and the immediate response to the attacks. The Commission is also charged with providing recommendations designed to guard against future attacks.

On July 22, 2004 the Commission released its public report, which is available for download from this site. The report is also available in bookstore editions from the Government Printing Office.

On August 21, 2004 the Commission released two staff monographs, available for download along with other staff statements on this site.

The ten members of the 9-11 Commission announce the creation of the [Executive Project](#).

The Commission closed on August 21, 2004.

### Your President's National Partnership for Reinventing Government

INDEX COMMENTS Archived in our PDF

SEARCH PHRASEY

**Test Your Retention IQ**  
REGO E-zine  
Access America  
Who's My Guy  
Accomplishments


AfterSchool  
BookSmiles  
Community  
E-Gov  
Federal Workplace  
Hammer Awards  
Healthy Few Communities  
High Impact Agencies  
Initiatives  
Jobs / Careers  
Participating  
Plain Language  
Registration  
SubCircles

**Celebrating Reinyer**  
"No executive branch reform is needed, perhaps 210 years - has enjoyed attention over such a brief for such a long period of in 1998. Here's what we

**1,378 Hammer Award Federal Workers for Government**  
"When the Vice President Administration's bold job reinventing government effort for Reinventing Government thought it was the career sended being and who was losing change."

[NPS Alumni Directory](#) (Home info)  
[GovExec.com](#)  
[Clinton Administration Awards List](#)


[www.hatpost.com](#), [Packin' Up, Not Giving!](#)



### COLUMBIA ACCIDENT INVESTIGATION BOARD

HOME | BOARD CHARTER | BOARD MEMBERS | NEWS | PHOTOS | EVENTS | CONTACT THE CAIR

On Feb. 1, 2003, Shuttle Columbia was lost during its return to Earth. Investigators have found the cause.



**Recently Released**

October 28, 2003 [Columbia Accident Investigation Board Releases Web 2.0 Final Report](#)  
The Columbia Accident Investigation Board today released Web 2.0 of the CAIB's Final Report. These volumes contain information that provides the supporting documentation for the main text of the Final Report contained in Vol. 1. [View](#)


October 28, 2003 [The CAIB Report, Volume 2.0](#)

October 24, 2003 [Columbia Accident Investigation Board 2: Release Web 2.0 of Final Report](#)  
The Columbia Accident Investigation Board will release Web 2.0 of the CAIB's Final Report at 10 a.m. on Tuesday, Oct. 28, 2003 on its website. [View](#)

August 24, 2003 [The CAIB Report, Volume 1](#)

**Events**

[Press Briefing](#)  
Tuesday, August 26  
11:00 a.m. EDT  
National Transportation Safety Board  
Conference Center  
420 L'Enfer Plaza, 5th  
Washington, D.C.



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

## WHAT IS THE CYBERCEMETERY?

- ✘ **online archive of websites** from U.S. government agencies or commissions that are no longer operating
- ✘ maintained by the University of North Texas Libraries
- ✘ freely accessible world-wide



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

## WHY CREATE THE CYBERCEMETERY?

- × At-Risk Information:

- + 1990's: U.S. government information moved online
- + much of it born-digital
- + often edited or removed without warning

- × Federal Depository Library Program (FDLP)



- × administered by the U.S. Government Printing Office (GPO)
- × mission: to provide free, permanent public access to government information
- × online information complicates this mission
- × University of North Texas is a federal depository library

## DEVELOPMENT

---

### × 1995

- + Government Printing Office (GPO) publishes report stating need to preserve electronic government publications



### × 1997

- + University of North Texas (UNT) talks to GPO about forming a partnership
- + UNT archives website of the Advisory Commission on Intergovernmental Relations

## DEVELOPMENT

---

### × 1999

- + UNT/GPO partnership is expanded
  - × permanent public access
  - × multiple government websites
  - × government agency or commission which is no longer operating (and/or has issued a final report)
- + the collection is named “CyberCemetery”
  - × due to its collection of websites from “dead” government agencies and commissions





## DEVELOPMENT

---

### × 2006

+ UNT/GPO partnership is expanded

- × Now includes the U.S. National Archives and Records Administration (NARA)



## ARCHIVING PROCESS

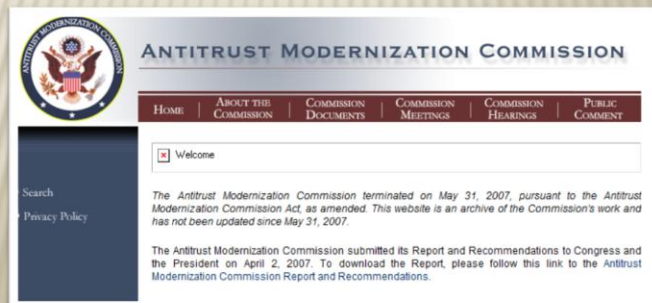
---

1. **Identify at-risk government agencies and commissions**
  - + read/listen to the news
  - + read government-related websites and blogs
  - + get referrals from other librarians
  - + contacted by GPO
  - + contacted directly by the agency/commission

# ARCHIVING PROCESS

## 2. Evaluate the website

- + must be an official government website
- + The agency or commission must:
  - × Be closing
  - × Have issued a final report
  - × Other indication that the website is at-risk



# ARCHIVING PROCESS

---

## 2. Evaluate the website (continued)

### + Questions for website administrator:

- + *What operating system was used to host this website?*
- + *What webserver software was used for the hosting of this website?*
- + *Are server side includes (ssi) used in this website?*
- + *Was this website static html or a dynamic site?*
  - + *If dynamic, what scripting languages were used for this website (php, perl, python)?*
  - + *Was a database used for this website?*
    - + *If so, what database was used for this website?*
    - + *What methods were used to connect to the database?*
- + *Is there streaming media associated with this website?*
- + *Are there proprietary content types used in this website?*
- + *Are there any comments you would like to add?*

## ARCHIVING PROCESS

---

### 3. Harvest the website

- + Past method: HTTrack
  - × <http://www.httrack.com/>
  
- + Future method: Heritrix
  - × <http://crawler.archive.org/>
  - × ARC files
  - × website in a single file: 100 – 600MB

## ARCHIVING PROCESS

---

### 3. Harvesting alternative: **Donated content**

- + directly receive files from agency or commission

- × Why donated content?

- \* If content cannot be accessed by harvesting
- \* flash video, large amounts of media

- × Why *not* donated content?

- \* Content could be altered
- \* Harvesting = exact copy of online published content

# ARCHIVING PROCESS

---

## 4. Link Checking

### + Automated:

- × Xenu Link Checker
- × <http://home.snafu.de/tilman/xenulink.html>
- × compare reports of original and archived sites

### + Manual:

- × manually navigate original and archived sites

## ARCHIVING PROCESS

---

### 5. Archive Preparation

- + add text “Archive”
  - × 8 point, Times New Roman font
  - × added to top/center of each page
- + manually disable contact links
  - × “mail to” links
  - × submit-able forms
- + (Heritrix will make these preparations unnecessary)



# ARCHIVING PROCESS

## 6. Load to UNT Server

- + Upload archived website
- + Add navigation
- + Notify GPO (or agency/commission) that archived version is live

The screenshot displays the CyberCemetery website interface. The top navigation bar includes the CyberCemetery logo and the University of North Texas Libraries. Below the navigation bar, there are three main content areas:

- Left Sidebar:** Contains links for CyberCemetery Home, Search the CyberCemetery, Browse the CyberCemetery, Related Resources, Contact Us, and Digital Collections Home.
- Center Panel:** Titled "Browse the CyberCemetery: Agencies by Date Expired". It features a list of agencies categorized by year (2007, 2006, 2004). Examples include the "Antitrust/Moderation Commission (May 2007)", "President's Advisory Panel on Federal Tax Returns (November 2006)", and "United States Commission on Ocean Policy (December 2004)".
- Right Panel:** Titled "Browse the CyberCemetery: Agencies by Branch of Government". It provides a list of agencies categorized by branch, such as Executive Branch, Legislative Branch, and Independent Commissions.

## TECHNICAL DETAILS

---

### ✘ Equipment

- + Four servers (three as backup)
- + Four node fail-over clustered configuration
- + SAN volume
- + 27.2GB of content on 40GB server

### ✘ Environment

- + Library basement
- + 38° Fahrenheit (3° Celsius)
- + 50% humidity



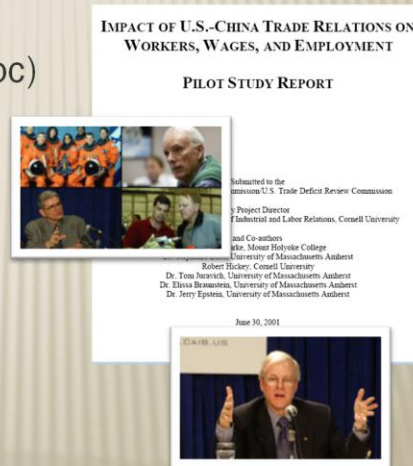
## TECHNICAL DETAILS

---

- × Backup
  - + full backups to magnetic tape
  - + performed each weekend
  - + shipped to offsite storage company
    - × Iron Mountain
    - × <http://www.ironmountain.com>

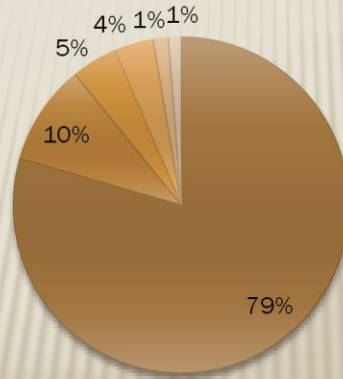
# TYPES OF CONTENT

- ✗ web files (HTML, XML)
- ✗ text documents (.txt, .pdf, .doc)
- ✗ spreadsheets & statistical information (.xls)
- ✗ presentations (.ppt)
- ✗ media files:
  - + images & photographs (.jpg, .gif, .png, tiff)
  - + audio (.mp3)
  - + video (.wm, .mov, .rp)



## WHERE ARE OUR USERS?

- United States
- Other/Unknown
- Continental Europe
- Asia
- Canada
- United Kingdom



## CONCLUSION

---

- ✘ provides **permanent public access**
- ✘ archive of government information
- ✘ **freely, globally available**
- ✘ 45 websites and growing
- ✘ **partnership** between:
  - + University of North Texas Libraries
  - + U.S. Government Printing Office
  - + National Archives and Records Administration

FOR FURTHER INFORMATION:

- × <http://govinfo.library.unt.edu>
- × <http://www.library.unt.edu/govinfo/digital-collections>

× **Starr Hoffman**

Librarian for Digital Collections

University of North Texas Libraries

[starr.hoffman@unt.edu](mailto:starr.hoffman@unt.edu)

<http://geekyartistlibrarian.blogspot.com>

---