PRESERVATION & DIGITAL COLLECTIONS

@ the University of North Texas

“Government information preservation: collections and collaboration”
ALA GODORT Program / ALA Annual 2017, Chicago
CyberCemetery

About This Collection

The CyberCemetery is an archive of government web sites that have ceased operation (usually web sites 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 web sites that cover topics supporting the university’s curriculum and particular program strengths.

Background

Foreseeing the potential preservation problems created by federal agencies’ ventures into electronic publishing, UNT became the second depository library in the nation to join the Federal Depository Library Program Content Partnerships Program. This program attempts to ensure permanent public access to electronic federal information. As a participant, the UNT Libraries, part of the Federal Depository Library Program (FDLP), were designated the host of the permanent online collection of the defunct Advisory Commission on Intergovernmental Relations (ACIR).
UNT DIGITAL LIBRARY

• 365,141 items
• 16.6 million files
• 118 Collections from 31 contributing partners
• 22.2 million total uses
  • 475,933 uses in the last 34 days (as of June 19, 2017)

UNT GOVERNMENT DOCUMENTS
(CONTRIBUTING PARTNER)

• 254,960 items*
• 13.2 million files*
• 11.4 million total uses*
• 132,436 uses in last month (as of June 19, 2017)

(statistics are combined totals for UNT Digital Library and Portal to Texas History)
WHY DIGITAL COLLECTIONS

Guiding Principles @ UNT

The UNT Libraries’ Digital Collections are maintained in accordance with the following guiding principles:

• Provide unified digital access to resources of value to the UNT Community
• Make access to these digital collections easy and transparent
• Sustainably manage the digital collections in perpetuity

SCOPING AND CREATING DIGITAL COLLECTIONS

A few examples from UNT and collaborative efforts
USDA Farmers' Bulletins

The United States Department of Agriculture Farmers' Bulletins were produced to disseminate information about agricultural topics. This collection includes bulletins published between the 1880's and the 1980's.
Congressional Research Service Reports

About This Collection

The Congressional Research Service (CRS) does not provide direct public access to its reports, requiring citizens to request them from their Members of Congress. Some Members, as well as several non-profit groups, have posted the reports on their web sites. This site is not affiliated with the Congressional Research Service, but aims to provide integrated, searchable access to many of the full-text CRS reports that have been available at a variety of different web sites since 1990.

The information on this site is compiled and made available as a public service by the Government Documents Department at the UNT Libraries. UNT does not make any warranty as to the accuracy, reliability, or completeness of the information and is not responsible for any errors or omissions or for results obtained from the use of the information. Distribution of the information does not constitute such a warranty. Use of the information is the sole responsibility of the user.
Technical Report Archive and Image Library (TRAIL)

This selection of materials from the Technical Report Archive and Image Library (TRAIL) includes hard-to-find reports published by various government agencies. The technical publications contain reports, images, and technical descriptions of research performed for U.S. government agencies. Topics range from mining, desalination, and radiation to broader physics, biology, and chemistry studies. Some reports include maps, foldouts, blueprints, and other oversize materials.
Presented by James Jacobs


**Archive Scope**

The End of Term Web Archive contains federal government websites (.gov, .mil, etc) in the Legislative, Executive, or Judicial branches of the government. Websites that were at risk of changing (i.e., whitehouse.gov) or disappearing altogether during government transitions were captured. Local or state government websites, or any other site not part of the federal government domain were out of scope.
EOT 2016 PARTNERS!

INTERNET ARCHIVE

LIBRARY OF CONGRESS

University of California
California Digital Library

NORTH TEXAS

Stanford LIBRARIES

GW Libraries

Federal Government Web Archiving Working Group

Slide courtesy of James R. Jacobs
GOALS OF THE END OF TERM PROJECT

- Work collaboratively to preserve public U.S. Government websites
- Document federal agencies’ presence on the web at the end of Presidential terms
- Enhance the existing research collections of the partner institutions
- Raise awareness about the need for preservation
- Engage with researchers, subject experts, journalists and the public
eot collaborative distribution of work

- IA: infrastructure, crawling, preservation, access, full-text search, metadata, research & support
- LC: crawling, preservation, data transfers
- UNT: nomination tool development, crawling, nomination mgmt, preservation, access
- GWU: Social media collection
- CDL: web portal, metadata
- Stanford: URL contributions, outreach, project mgmt
- GPO: URL nomination, outreach
- Others: URLs, education, coordination
eot 2016 results

- ~300 TB data total
  - ~110 TB web crawls + ~130 TB of gov ftp site archiving
- 310,000,000 web URLs + 12,000,000 ftp files
- 9,000+ social media accounts (scrape of gov SM registry API) 44% FB, 37% TW, 10% YT
- ~190K total domain, subdomains, gov/non-gov sites and volunteer nominators ROCK!!
  - 2008: 457 from 26 nominators
  - 2012: 1476 from 31 nominators
  - 2016: 15,000+ from 400+ nominators (via UNT form)
  - Plus: Over 100,000 from DataRescue/EDGI events/tools
NEXT STEPS
• What is PEGI
  • A coordinated effort of libraries, federal agencies, and non-profit organizations to raise awareness around and address national concerns regarding preservation of electronic government information

• PEGI Activities to Date
  • 2 meetings aligned with the Coalition for Networked Information (CNI)
  • Organization of a Steering Committee
    • Conference calls and in-person meetings of the steering committee have directed and informed PEGI efforts
  • Interviews with federal information producers
  • Interviews with representatives from NARA, GPO and LC
Preservation of Electronic Government Information (PEGI)

- PEGI Next Steps
  - IMLS National Forum Grant Submission
  - Additional interviews with information producers, researchers and other stakeholders
  - Provide a report that defines stakeholder priorities and a common agenda, a plan of collaborative action and a means for stakeholders to communicate and continue efforts toward a common agenda
  - https://www.crl.edu/preservation-electronic-government-information-pegi
defining the “government web presence”
AFFILIATED EFFORTS

- US government FTP sites
- DataRefuge
- Climate Mirror: an open project to mirror public climate datasets
- Azimuth Climate Data Backup Project

Slide courtesy of James R. Jacobs