


U.S. GOVERNMENT PRINTING OFFICE | KEEPING AMERICA INFORMED

Meeting the Mission: Permanent Public Access in the Digital Age

GPO's strategies for managing and preserving government information

24th Annual Interagency Depository Seminar
August 4, 2011
David Walls, Preservation Librarian

our strategic vision in progress



U.S. GOVERNMENT PRINTING OFFICE | KEEPING AMERICA INFORMED

Our Information Landscape:

- Approximately 97% of current US Government Information is born-digital.

Subscriptions to the *Federal Register* in 1994, >20,000.
In 2010, <3000.
Congressional Record in 1994, >20,000. In 2010, <5,000.

- 59% of adults get their information from the mobile web on smart phones and tablet computers.

our strategic vision in progress

1

Our Information Landscape:

- We have a 150 years of collective experience managing and preserving paper-based information.
- We have little more than 15 years of collective experience managing and preserving digital information.

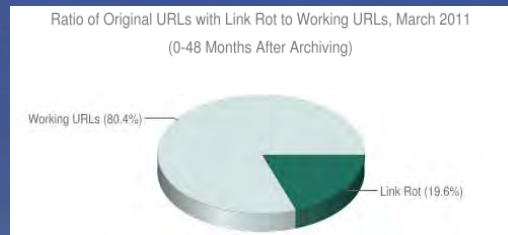
Mission: Permanent Public Access

Preservation may be defined as the development and implementation of strategic initiatives designed to maintain permanent access to information.

Our Information Landscape:

Web-based information is ephemeral!

“Link Rot” & Legal Resources on the Web:
A 2011 Analysis” www.legalinfoarchive.org



Strategic Initiatives:

1. Web Harvesting



contact
the search with google

About:

mission
who we are
member services
for members
faq
working groups
press releases

publications:

reporting
membership and
website

software:

hawaii
preserv3

Mission

The mission of the International Internet Preservation Consortium (IIPC) is to acquire, preserve and make accessible knowledge and information from the Internet for future generations everywhere, promoting global exchange and international relations.

Goals

To achieve its mission, the IIPC is working to accomplish the following goals:

- To enable the collection of a rich body of Internet content from around the world to be preserved in a way that it can be archived, secured and accessed over time.
- To foster the development and use of common tools, techniques and standards that enable the creation of international archives.
- To encourage and support national libraries everywhere to address Internet archiving and preservation.

Charter

In July 2002, the IIPC was formally chartered at the Bibliothèque nationale de France with 12 participating institutions. The members agreed jointly to fund and participate in projects and working groups to accomplish the goals of the IIPC.

The initial agreement was in effect for three years, and membership was limited to charter institutions. The IIPC is now open to libraries, archives, museums and cultural heritage institutions everywhere and welcomes inquiries for [membership](#).



http://www.w3.org/2004/01/IIPC/Charter-and-Access-Statement/Charter

CyberCemetery

University of North Texas | History

CyberCemetery Home

Search the CyberCemetery

Browse the CyberCemetery

Related Resources

Contact Us

Digital Collections Home

UNT Libraries
Government Documents Dept
1150 Union Circle #30370
Ft. Worth, TX 76107-3037
Phone: 817.257.3037
Email: libref@unt.edu
Library Fax: 817.257.2538

What is the CyberCemetery?

The **CyberCemetery** is an archive of government websites that have ceased operation (usually websites of defunct government agencies and commissions that have ceased a their work). This collection includes a variety of topic materials of the broad nature of government information. In particular, this collection features websites that provide support for the university's curriculum and academic program objectives.

Who created and maintains the CyberCemetery?

The **University of North Texas Libraries** and the **U.S. Government Printing Office**, as part of the Federal Depository Library Program, created a partnership to provide permanent public access to the Web sites and publications of defunct U.S. government agencies and commissions. This collection was named the "CyberCemetery" to help users of the site.

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

Strategic Initiatives & Technology Tools:

1. Harvest web content
2. Develop a digital repository

The Federal Digital System, FDsys



A Digital Repository:

GPO's Federal Digital System, FDsys, a digital content management and repository for US government publications.

- Replaces "GPO Access"
- Extensible system architecture
- Extensive Metadata for searching
- Cryptographic hash authentication of all content. Chain of custody PKI on selected content.
- Designed around OAIS model to be a "trustworthy" digital repository

What is FDsys?

- FDsys is a *Content Management System*
 - FDsys securely controls digital content throughout its lifecycle to ensure content integrity and authenticity
- FDsys is an *Advanced Search Engine*
 - FDsys combines extensive metadata creation with modern search technology
- FDsys is a *Preservation Repository*
 - FDsys follows archival system standards to ensure the long-term preservation and access of digital content.
 - Has been self-audited according to TRAC metrics.

FDsys Content:

1. **Deposited Content:**
Federal Agencies deposit digital publications in FDsys.
2. **Harvested Content:**
GPO harvests and archives publications within the scope of the FDLP.
3. **Digitized Content:**
Selected FDLP legacy publications

The screenshot shows a web browser window displaying the FDsys page for SUBCHAPTER A—GENERAL, PART 400—BASIS AND SCOPE. The page content includes:

SUBCHAPTER A—GENERAL

PART 400—BASIS AND SCOPE

Sec.
400.1 Basis.
400.2 Scope.

Airways: 49 U.S.C. 70101-70101.
Statute: Docket No. 4810-53 FR 11023, Apr. 4, 1985, unless otherwise noted.

§ 400.1 Basis.
The basis for the regulations in this chapter is the Commercial Space Launch Act of 1984, and applicable treaties and international agreements to which the United States is party.

§ 400.2 Scope.
These regulations set forth the procedures and requirements applicable to the authorization and supervision under 49 U.S.C. Subtitle IX, chapter 701, of commercial space transportation activities conducted in the United States or by a U.S. citizen. The regulations in this chapter do not apply to space-related activities at quarters, 800 Independence Avenue, SW., Room 331, Washington, DC 20501.
(AM09, 49-3, 68 FR 32395, June 18, 2003)

§ 401.3 The Associate Administrator for Commercial Space Transportation.
The Office is headed by an Associate Administrator to exercise the Secretary's authority to license or permit and otherwise regulate commercial space transportation and to discharge the Secretary's responsibility to encourage, facilitate, and promote commercial space transportation by the United States private sector.
(Doc. No. FAA-2006-24197, 71 FR 1706, Apr. 6, 2007)

§ 401.5 Definitions.
As used in this chapter—
Act means 49 U.S.C. Subtitle IX, Commercial Space Transportation, ch. 701—Commercial Space Launch Activities, 49 U.S.C. 70101-70121.
Associate Administrator means the As-

Strategic Initiatives:

1. Harvest web content
2. Develop a digital repository
3. Develop partnerships with other agencies

Partnerships:

- Identify maps, GIS data, databases, preservation file format copies, and other web publications not capable of being harvested and develop partnerships with agencies creating content.
- Reach out to agencies to encourage deposit of agency content with GPO.
- Maintain networking relationships with agencies focused on sharing best practices.
- Leverage LC partnership for web harvesting and digitization.
- GPO is a NARA electronic affiliate.

Strategic Initiatives:

1. Harvest web content
2. Develop a digital repository
3. Develop partnerships with other agencies
4. Participate in developing digitization guidelines and best practices

Privacy | Partnerships | Contact Us

**Federal Agencies
Digitization Guidelines Initiative**

[Link Collection](#)
Submit comments or guidelines to our digitization efforts in general.

Home
About
Guidelines
Resources
Review Process
News & Events

About This Initiative

Started in 2007, this is a collaborative effort for Federal agencies to derive common guidelines, methods, and practices for digitizing historical content. As part of this, two working groups are studying content specific to two major areas: Still Image and Audio Visual.

[Learn more about the initiative](#) >

Still Image Working Group

This group is involved in a cooperative effort to develop common digitization guidelines for still image materials.

Audio Visual Working Group

The goal for this working group is to identify, establish, and disseminate information about standards and practices for the digital recording of audio-visual materials.

Digitization Guidelines

Technical Guidelines for the Still Image Digitization of Cultural Heritage Materials
Updated by Working Group, August 24, 2010

Creation of Reader Image Master Files represents shared best practices followed by agencies participating in the Federal Agencies Digitization Guidelines Initiative (FADGI) Still Image Working Group for digitizing cultural heritage material.

MXF Application Specification for Moving Image and other Audio-Visual Content
Draft under development

The MXF Application Specification for Archiving and Preservation is a detailed specification for a file "wrapper" intended to serve as an interlocking program in which an archive is making a digital file copy of a videotape or motion picture film, as well as to serve as a container for born-digital content when the native encoded system it is being outside the retention for at least the medium term.

Audio Digitization System Performance
Concept under development

Pertaining to the performance testing of audio digitization systems, the next emerging component of a draft report. When ready, the activity will parallel efforts by the Still Image Working Group to develop metrics and methods for the performance testing of archival test cameras. Interested parties are invited to comment.

TIFF Image Metadata
Approved by Working Group, February 10, 2009

A new recommended guideline accompanying a richer metadata set in a more flexible format, with wider file format support, is expected to be developed by the Proposed Metadata Sub-group and posted at a

News & Events

[Evaluating Digitization and Digitization Equipment](#) (PDF, 23 MB) | August 16, 2010

[Information Standards Quarterly article](#) (PDF, 528 KB) | Spring 2010

[Federal Computer Week article of February 5, 2010](#)

[See all News & Events](#) >

Resources

The following are provided as resources for the digitization of both audio-visual and still image digital content:

- [Digital Assets Checklist](#)
- [Evaluating Still Image Digitization and Digitization Equipment](#) (PDF, 23 MB)
- [TIFF Image Metadata in Audio Visual Preservation Recommendation](#) (PDF, 528 KB)

[View all Resources](#) >

our strategic vision in progress

17

9

Strategic Initiatives:

1. Harvest web content
2. Develop a digital repository
3. Develop partnerships with other agencies
4. Participate in developing digitization guidelines and best practices
5. Facilitate the digitization of selected legacy publications

Preservation Problems with legacy collections

- Pre-1986 content is fragile due to acidic decay.
- Content is vulnerable to fire, flood, pests, environmental disasters, people who are too lazy to photocopy, and other acts of vandalism
- Primary Source and other special collection materials are unique and vulnerable
- No way to backup content to maintain continuity of access
- Collections stored out of sight become collections stored out of mind
- Divide between paper and digital content widening



<http://registry.fdlp.gov/>



Strategic Initiatives:

1. Harvest web content
2. Develop a digital repository
3. Develop partnerships with other agencies
4. Participate in developing digitization guidelines and best practices
5. Facilitate the digitization of selected legacy publications
6. Network with community

FDLP, Federal Agencies, Congress



Permanent Public Access in the Digital Age

1. Harvest web content
2. Develop a digital repository
3. Develop partnerships with other agencies
4. Participate in developing digitization guidelines and best practices
5. Facilitate the digitization of selected legacy publications
6. Network with community

Ongoing Issues:

- Harvesting interactive databases and streaming web content
- PII (personally identifiable information)
- Authentication of all content
- Modeling costs, potential user services, and ingest requirements for FDsys

GPO

David Walls
Preservation Librarian
dwalls@gpo.gov