Current Issues in Digital Stewardship

2013 Depository Library Council Meeting and Federal Depository Library Conference
July 31, 2013

Butch Lazorchak
Digital Archivist
Library of Congress
National Digital Information Infrastructure and Preservation Program
wlaz@loc.gov
facebook.com/digitalpreservation
@ndiipp

Outline

- Background on the National Digital Information Infrastructure and Preservation Program at the Library of Congress, including:
  - Viewshare
  - Federal Agencies Digitization Guidelines Initiative
  - Personal Digital Archiving
  - Community Building and the National Digital Stewardship Alliance
- Digital Stewardship issues of special concern to the Depository Library Community:
  - File Formats for Long-term stewardship
  - Preserving Digital Legislative Information
  - Web Archiving
  - Digital Geospatial Data (very briefly!)

Repository of the Past?

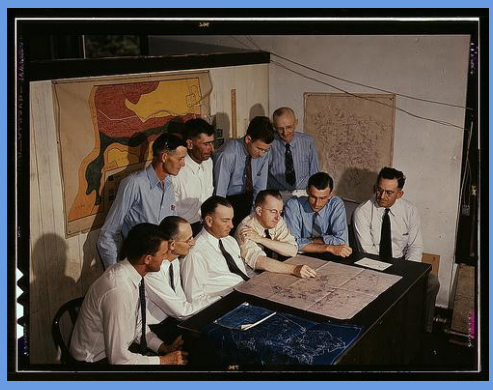
Predictor of the future?

National Digital Information Infrastructure and Preservation Program

MISSION: Ensure access over time to a rich body of digital content through establishment of a national network of partners committed to selecting, collecting and preserving at-risk digital information.

http://www.digitalpreservation.gov

NDIIPP Focus Areas

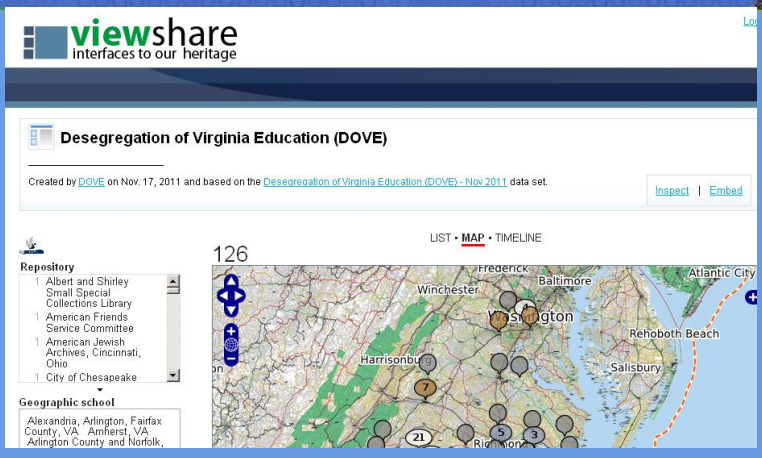


- Content
- Partnerships: Government, Industry, Academia
- Technical Infrastructure
- Sustainability
- Outreach and Education

Some Initiatives

- Viewshare
- Federal Agencies Digitization Guidelines Initiative
- Personal Digital Archiving
- Community Building and the National Digital Stewardship Alliance
- Outreach and Communication

Technical Infrastructure: Viewshare



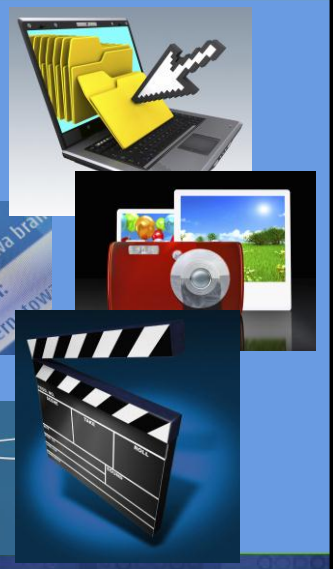
Viewshare is a free platform for generating and customizing views, (interactive maps, timelines, facets, tag clouds) that allow users to experience your digital collections. http://viewshare.org/

Federal Agencies Digitization Guidelines Initiative (FADGI)

- Started in 2007, FADGI is a collaborative effort by federal agencies to define common guidelines, methods, and practices for digitizing historical content. As part of this, two working groups are studying issues specific to two major areas, Still Image and Audio Visual.
- Digital Imaging Standards
- Format Considerations in Audio Visual Preservation Reformatting
- http://www.digitizationguidelines.gov/

Personal Digital Archiving

Guidance to the "general public" on saving their digital stuff: documents, photos, music, video, email, websites etc.



- Public Events
- Further How-to's and tutorials
- Preservation Week events such as "Personal Archiving Day"
- Co-Hosting Personal Digital Archiving 2014 in Indianapolis

http://www.digitalpreservation.gov/personalarchiving/

Community Building

- Preserving.exe: Toward a National Strategy for Preserving Software 2013
- Personal Digital Archiving 2013
- Designing Storage Architectures for Preservation Collections
- Science at Risk: Toward a National Strategy for Preserving Online Science 2012
- Upcoming Meetings on History and Social Science
- http://www.digitalpreservation.gov/meetings/

National Digital Stewardship Alliance

- http://www.digitalpreservation.gov/ndsa
- https://twitter.com/NDSA2
- http://www.loc.gov/extranet/wiki/osi/ndiip/ndsa/index.php?title=Main_Page

National Digital Stewardship Alliance

NDSA

• An initiative of the National Digital Information Infrastructure and Preservation Program (NDIIPP)

• A collaborative effort among government agencies, educational institutions, non-profit organizations, and business entities to preserve a distributed national digital collection for the benefit of citizens now and in the future.

• Community-driven, action-oriented

NDSA Membership

• Over 150 member organizations have joined the NDSA since it was founded in July 2010.

• Examples: Federal Library and Information Center Committee (FLICC), U.S. Fish and Wildlife Service, Northeast Document Conservation Center, National Park Service, New York Public Library, National Library of Medicine, Thomson Reuters, Academy of Motion Picture Arts and Sciences, California Digital Library, Harvard University, NARA, OCLC, Public Broadcasting Service, many more.

• Membership is now open to any interested U.S.-based organizations. Very lightweight membership agreement!

Benefits of Participation

- Have a Voice
- Share knowledge
- Drive action
- Learn from Experts
- Impact change

NDSA Governance

Eight Member Coordinating Committee drawn from all sectors.

Members commit to participating in one of the five working groups:

- Content
- Infrastructure
- Innovation
- Outreach
- Standards and Practices

NDSA Working Group Activities

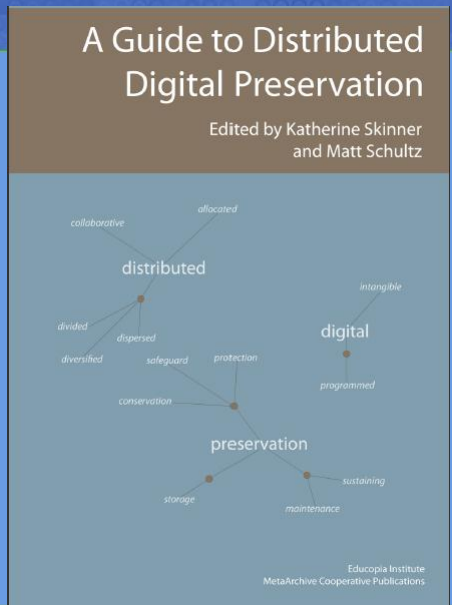
- Preservation Levels
- PDF/A white paper
- Cloud Presentations
- Innovation Awards
- Geospatial Content Appraisal white paper
- Digital Preservation in a Box
- Kickstarter Curator Page
- Standards and Best Practices Knowledge Base

Outreach

- http://www.digitalpreservation.gov
- Monthly newsletter (19,100+ subscribers)
- Twitter: twitter.com/ndiipp or @ndiipp (13,700+ followers)
- Facebook: facebook.com/digitalpreservation (5,300+ likes)
- Blog: http://blogs.loc.gov/digitalpreservation/
- Videos
- Podcasts
- National Book Festival
- Printed brochures
- Smithsonian preservation events
- Preservation Week/Day Events

And Wait! There's More!

http://ptup.editme.com/files/2007-SJ-Monterey-Tour/TheWholeEnchilada.jpg



- Distributed Architectures
- Technical Considerations
- Organizational Considerations
- Content Selection, Preparation, and Management
- Content Ingest, Monitoring, and Recovery
- Network Administration

http://www.metaarchive.org/GDDP

Digital Stewardship issues of special concern to depository libraries

- File Formats for Long-term stewardship/PDF-A
- Web Archiving
- Preserving Digital Legislative Information
- Digital Geospatial Data
- Distributed Digital Preservation/Cloud Computing

File Format Sustainability

- Library of Congress Sustainability of Digital Formats: Planning for Library of Congress Collections
- PDF/A

http://www.digitalpreservation.gov/formats/index.shtml

What is PDF/A?

- Some useful features of PDF are incompatible with the demands of long-term preservation
- Preserve the “static visual appearance”
- Self-contained
- Must include embedded fonts
- Must include device-independent color
- Must include XMP metadata
- May not include: Encryption* (digital signatures ok in V2,3), LZW compression, embedded files* (now OK in V3), external content references, javascript, etc.

PDF/A-1, -2, -3

- “Real” name of first version is ISO 19005-1:2005 Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1)
- “Real” name of the second version is ISO 19005-2:2011 Document management -- Electronic document file format for long-term preservation -- Part 2: Use of ISO 32000-1 (PDF/A-2)
- “Real” name of the third version: ISO 19005-3.2 Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)

PDF/A Resources

- PDF Associaton PDF/A Competence Center: http://www.pdfa.org/competence-centers/pdfa-competence-center/
- LC Sustainability Site PDF/A-1 Page: http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml
- D-Lib Magazine: PDF/A: A Viable Addition to the Preservation toolkit http://dlib.org/dlib/november10/noonan/11noonan.html

Preserving Digital Legislative Information

Model Technological and Social Architecture for the Preservation of State Government Digital Information (MTSA)

- Authentication Resources
- Authentication White Papers
- Best Practice Principles for Opening Up Government Information
- Business Case for Digital Preservation
- Cloud Computing
- Digital Audio Video White Paper and Resources
- Government Data Mashups White Paper
- Legislative Metadata
- Legislative History Resources
- Options for Improving Access to Legislative Records
- Project Podcast
- Preservation Options
- Record Inventory
- Records Retention Policies for Selected Legislative Records
- Retrospective Digitization White Paper and Resources
- Survey of Partner's Legislative Records on the Web
- Web Archiving and Evaluation
- Web Content Accessibility
- XML Basics
- XML Native Database White Paper
- XML Usage Survey

UELMA

- Uniform Electronic Legal Material Act
  http://www.uniformlaws.org/Act.aspx?title=Electronic%20Legal%20Material%20Act

•If an official publisher publishes legal material only in an electronic record, the publisher shall designate the electronic record as official

•An official publisher of legal material in an electronic record that is designated as official under shall authenticate the record.

•To authenticate an electronic record, the publisher shall provide a method for a user to determine that the record received by the user from the publisher is unaltered from the official record published by the publisher.

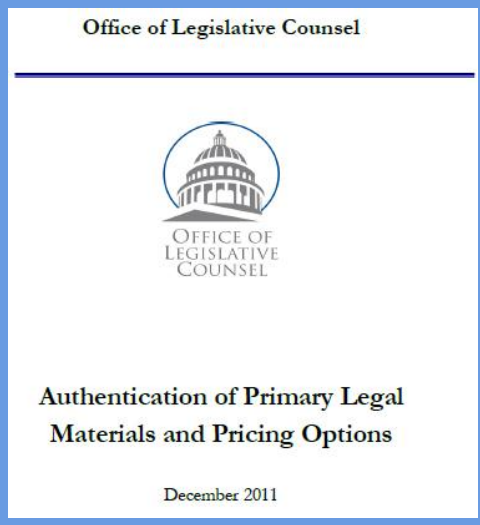
•An official publisher of legal material in an electronic record that is or was designated as official under Section 4 shall provide for the preservation and security of the record in an electronic form or a form that is not electronic.

Authentication: Documents, Not People

- AALL State-by-State Report on Permanent Public Access to Electronic Government Information (2003)
  - http://www.aallnet.org/Archived/Government-Relations/Issue-Briefs-and-Reports/2003/ppareport.html
- AALL State-By-State Report on Authentication of Online Legal Resources (2007)
  - http://www.aallnet.org/aallwash/authen_rprt/AuthenFinalReport.pdf
- GPO Authentication as part of FDSys (2011)
  - http://www.gpo.gov/pdfs/authentication/authenticationwhitepaper2011.pdf

Authentication of Primary Legal Materials and Pricing Options Paper

- Possible Authentication Methods: Secure Web sites; Document hashes; Digital Signatures; Public Key Infrastructures; more
- http://www.mnhs.org/preserve/records/legislativerecords/docs_pdfs/CA_Authentication_WhitePaper_Dec2011.pdf


Office of Legislative Counsel

OFFICE OF LEGISLATIVE COUNSEL

Authentication of Primary Legal Materials and Pricing Options

December 2011


Uniform Law Commission

The National Conference of Commissioners on Uniform State Laws

Contact Us: 312.450.6600 Login

Home | Acts | Committees | Legislation | Meetings | News | About ULC | Search

Acts

Electronic Legal Material Act

Legislative Tracking

2013 Introductions & Enactments

State	Bill	Sponsor	Status
Connecticut	SB 235		Enacted
District of Columbia	20-221	Wells	Introduced
Hawaii	SB 32	Hee	Enacted
Illinois	SB 1941	Mulroe	Introduced
Massachusetts	HB 38		Introduced
Minnesota	HF 278	Hortman	Enacted
Missouri	SB 478	Lamping	Introduced
Nevada	SB 105		Enacted
North Dakota	HB 1129		Enacted
Oregon	HB 2944	Davis	Enacted
Pennsylvania	SB 601	Scarnati	Introduced
Rhode Island	HB 5850	Hearn	Introduced

Act History

Origin: Completed by Uniform Law Commissioners in 2011.
Committee: Electronic Legal Material Act
Resources: Draft Materials

Enactment Status Map

= Enacted = Introduced this Year

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION
WASHINGTON, D.C. 20408

August 24, 2012

M-12-18

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES AND INDEPENDENT AGENCIES

FROM: Jeffrey D. Zients
Acting Director
Office of Management and Budget

David S. Ferriero
Archivist of the United States
National Archives and Records Administration

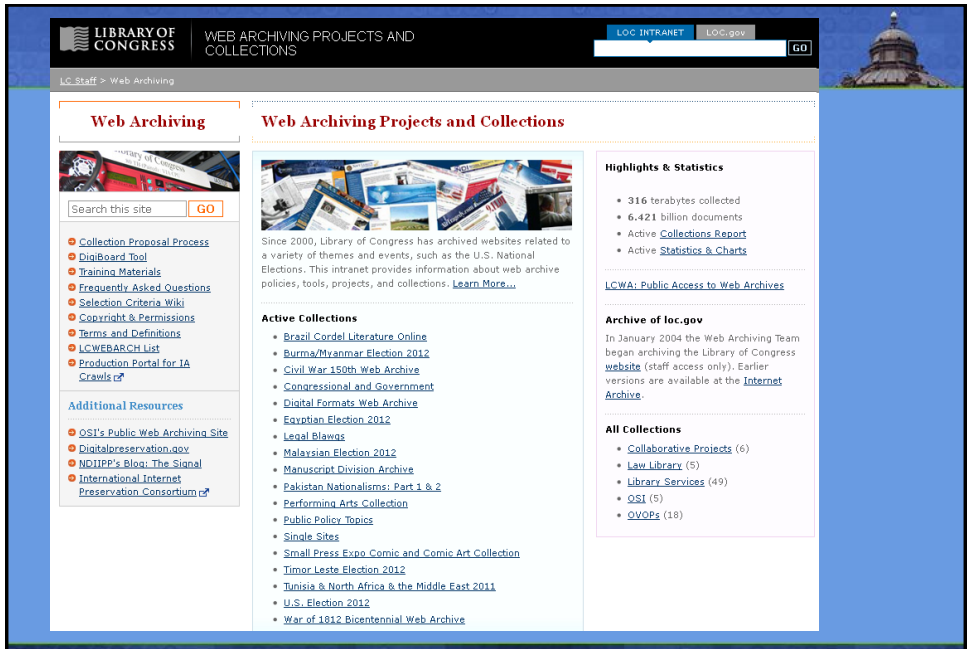
SUBJECT: Managing Government Records Directive

Legislative Preservation Resources

- Preserving Digital Legislative Information: Wrapping Up the MTSA Project
  - http://go.usa.gov/ylr
- NARA Records Express Blog
  - http://blogs.archives.gov/records-express/
- Center for Archival Resources On Legislatures (CAROL)
  - http://www.mnhs.org/preserve/records/legislativerecords/carol/index.htm
- Drafting Legislation Using XML at the U.S. House of Representatives
  - http://xml.house.gov/drafting.htm
- Authentication at the Government Printing Office
  - http://www.gpo.gov/authentication/
- Document Naming Conventions: To accompany the House standards for Electronic Posting of House and Committee Documents and Data
  - http://cha.house.gov/sites/republicans.cha.house.gov/files/documents/committee_docs/naming_conventions_v_1_2.pdf

Web Archiving

- LC web archiving activity
- Terminology
- Infrastructure
- Memento



http://www.loc.gov/webarchiving/

Web Archive Terminology

- **Seed -** the Nominated URL, or starting point for the crawler
- **Seed List** – List of Seed URLs fed to the crawler for harvesting
- **Scoping** – Instructions the team gives to the crawler to include other URLs associated with the seed (for example, crawl of whitehouse.gov would include facebook.com/whitehouse)
- **Crawler**– Software which downloads file, extracts URLs, adds URLs to list and repeats
- **Frequency** – How often a seed is crawled (one time, once a week, once a monthly, twice a year, or annually)

Web Archiving Infrastructure

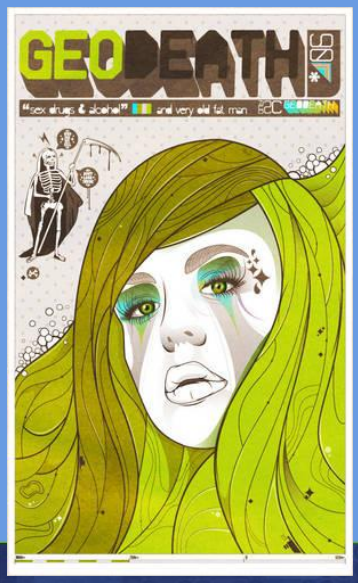
- Crawler: Heritrix
  - https://webarchive.jira.com/wiki/display/Heritrix/Heritrix
- Access: Wayback Machine
  - http://archive-access.sourceforge.net/projects/wayback/
- File format: WARC (ISO Standard)
  - http://bibnum.bnf.fr/WARC/
- Information Resource: International Internet Preservation Consortium (IIPC)
  - http://netpreserve.org

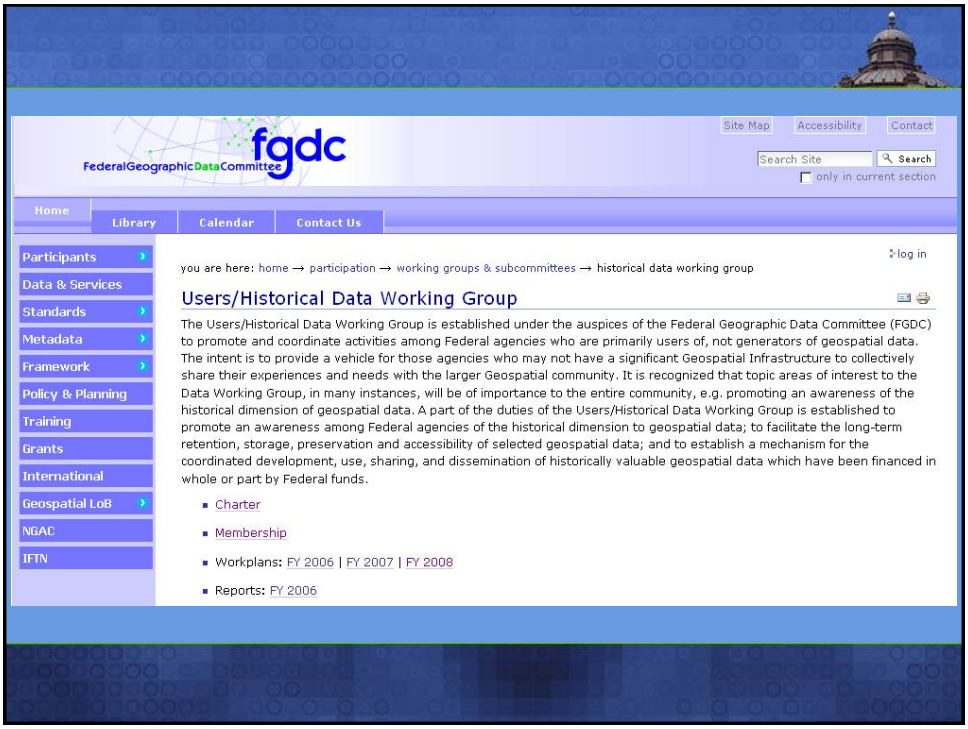
Memento

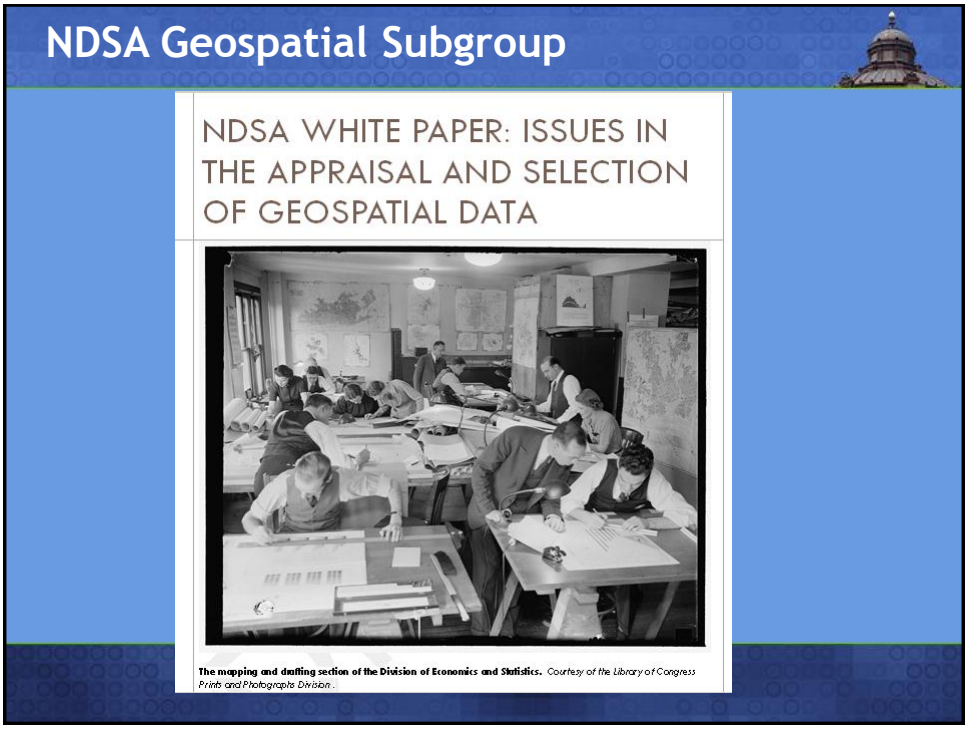
http://www.mementoweb.org/

What are the Special Risks to Geospatial Information?

- Unique geospatial data formats
- Spatial database complexity
- Fragility and uncertainty surrounding digital cartographic representation
- Issues related to time-versioned content
- Metadata unavailability or inconsistency
- No generally supported content packaging design for complex geospatial data




fgdc
FederalGeographicDataCommittee
Site Map
Accessibility
Contact
Search Site
Search
only in current section
Home
Library
Calendar
Contact Us
log in
Participants
Data & Services
Standards
Metadata
Framework
Policy & Planning
Training
Grants
International
Geospatial LoB
NGAC
IFTN
you are here: home → participation → working groups & subcommittees → historical data working group
Users/Historical Data Working Group
The Users/Historical Data Working Group is established under the auspices of the Federal Geographic Data Committee (FGDC) to promote and coordinate activities among Federal agencies who are primarily users of, not generators of geospatial data. The intent is to provide a vehicle for those agencies who may not have a significant Geospatial Infrastructure to collectively share their experiences and needs with the larger Geospatial community. It is recognized that topic areas of interest to the Data Working Group, in many instances, will be of importance to the entire community, e.g. promoting an awareness of the historical dimension of geospatial data. A part of the duties of the Users/Historical Data Working Group is established to promote an awareness among Federal agencies of the historical dimension to geospatial data; to facilitate the long-term retention, storage, preservation and accessibility of selected geospatial data; and to establish a mechanism for the coordinated development, use, sharing, and dissemination of historically valuable geospatial data which have been financed in whole or part by Federal funds.
Charter
Membership
Workplans: FY 2006 | FY 2007 | FY 2008
Reports: FY 2006


NDSA Geospatial Subgroup
NDSA WHITE PAPER: ISSUES IN THE APPRAISAL AND SELECTION OF GEOSPATIAL DATA
The mapping and drafting section of the Division of Economics and Statistics. Courtesy of the Library of Congress Prints and Photographs Division.

Geopreservation.org

Key Takeaways

- NDIIPP and the NDSA are trusted resources for digital stewardship information
- Digital Stewardship is multi-faceted, challenging and complex, but achievable
- We need your help to spread the word

http://shindigzparty.files.wordpress.com/2008/03/4s058a.jpg

Thanks!

Butch Lazorchak
Digital Archivist
Library of Congress
wlaz@loc.gov

http://imagecache2.allposters.com/images/pic/Matted_Prints/mp_814104_b~Man-on-Phone-Thanks-Posters.jpg