PEGI PROJECT

Developing a shared national agenda for the preservation of electronic government information

Martin Halbert (UNC Greensboro), Marie Concannon (University of Missouri) and PEGI Steering Committee Members

Preserving Electronic Government Information Project Webinar
Monday, 3/19/18
Librarians and other information professionals from University of North Texas, UNC Greensboro, Arizona State University, the University of Missouri, Yale, Stanford, the Center for Research Libraries, and the Educopia Institute are undertaking a multi-year project to address national concerns and develop a shared national agenda regarding the preservation of electronic government information (PEGI) by cultural memory organizations for long term use by the citizens of the USA.
Webinar Agenda

● Part 1 - Overview of the PEGI project

● Part 2 - Group discussion of three questions concerning your thoughts regarding preservation of electronic government information and ways of addressing the issue

● Part 3 - Invitation to participate in PEGI Project next steps
Presenter: Martin Halbert

- University of North Carolina at Greensboro Dean of Libraries
- President of the Educopia Institute and founder of the MetaArchive Cooperative during the National Digital Infrastructure and Information Preservation Program (NDIIPP)
- Previously worked as:
  - Dean of Libraries at the University of North Texas
  - Director for Digital Innovations at Emory University
Audience today

- Review of who’s in the webinar today
- Any observations on the categories of participants today
Part 1:

Overview of the PEGI Project, motivations and precursors
Prior work that led to the PEGI Project

● Some members of the Federal Depository Library Program have for many years been concerned that virtually all the attention concerning preservation efforts in the FDLP was being devoted to print materials.

● Noted that the vastly increasing amounts of electronic information produced by the Federal Government, or created with federal funding, was effectively being ignored by many members of the otherwise extremely conscientious FDLP system.

● There have been several efforts in the past to address this issue; PEGI builds on these efforts with the intent to broaden the conversation and address more pointed recent developments.
Precursors

- The Gateway Project: developed during the early days of GPO Access to maximize free public availability of the resources on GPO Access through federal depository library portals, project ended in 2000
- FDLP Content Partnerships Program: created in 1997 to address concerns about the need for permanent public access to electronic federal information, especially websites of commissions and agencies that concluded their work, became known as the CyberCemetery
- End of Term Crawl Project: has captured and preserved U.S. Government websites at the end of presidential administrations, in 2008 and 2012, and the 2016 through web-archiving crawls, and preserves federal government websites (.gov, .mil, .org)
- FIPNet: GPO initiative to expand FDLP collections and cataloging & indexing content for the National Bibliographic Records Inventory
The 2 x James Jacobs Efforts

- James (Jim) A. Jacobs (UC San Diego) and James R. Jacobs (Stanford University)
- Free Government Information: “a place for initiating dialogue and building consensus among the various players (libraries, government agencies, non-profit organizations, researchers, journalists, etc.) who have a stake in the preservation of and perpetual free access to government information” https://freegovinfo.info
- Both have published extensively and advocated tirelessly for preservation and free access to government information, especially electronic forms
Affiliated Archives of NARA

- U.S. Government Printing (now Publishing) Office (GPO)
- Library of Congress
- Oklahoma Historical Society
- Pennsylvania State Archives
- New Mexico State Records Center and Archives
- University of North Texas Libraries
- United States Military Academy
- United States Naval Academy
- Yellowstone National Park Archives

Source: https://www.archives.gov/locations/affiliated-archives
So What’s the big deal if we have all these efforts already? Why worry about the possibility of electronic government information loss?

- This is the U.S. Government right?
- All that information is always going to be around.
- Surely it’ll all be safe and secure.
404 ERROR!

Sorry, the page you tried to access cannot be found.

You may have typed the address incorrectly or you may have used an outdated link. If you found a broken link, please report it here.

Feedback?
Biggest Causes of Loss of Digital Information

- Operator error
- External Attack
- Insider Attack
- Economic Failure
- Organization Failure

Source: https://freegovinfo.info/node/4000
2016 PEGI Project Summits

● April Meeting (San Antonio, TX)
  ○ 2 day summit attended by 28 individuals representing 21 different organizations (universities, federal agencies, and nonprofits)
  ○ Articulated and analyzed the PEGI core issues
  ○ URL: https://digital.library.unt.edu/ark:/67531/metadc826639/

● December Meeting (Washington, D.C.)
  ○ Similar range of attendees further explored the PEGI core issues and developed the successful IMLS grant application
Core Issues (from 2016 report)

- In the pre-digital production era, a clear workflow more-or-less successfully accounted for the preservation of most government records and information.
- However, today, most government information is produced and disseminated digitally.
- This workflow is not nearly as well-understood and managed in a digital environment, and the sheer number and types of “electronic publications” has exploded.
- Individual federal agencies now have the ability to quickly and easily publish their work themselves, without involving the GPO or NARA.
- If a work is not categorized officially as a “record” or a “report” or a “publication,” regulatory authority does not require agencies to maintain content themselves, nor to provide it to the GPO or NARA for ongoing care.
- The goal of the first Summit was to gather the stakeholder groups involved in government information preservation to explore how best to ensure the longevity, accuracy, and accessibility of digital U.S. government information for an informed citizenry.

Source: https://digital.library.unt.edu/ark:/67531/metadc826639/
In Late 2016 the Issues Gained New Prominence

- The concerns over preserving electronic government information have now exploded
- Sudden advent of many grassroots efforts in the first months of 2017
  - DataRescues, Libraries+Network, EDGI, etc.
- The need now is no longer raising general awareness, the need now is for documenting:
  - The perspectives on this issue from the various concerned national stakeholder groups
  - Developing a shared agenda for addressing these concerns
**PEGI Purpose Statement**

Preservation of Electronic Government Information (PEGI) is a multi-year initiative to address national concerns and develop a shared national agenda regarding the preservation of electronic government information (PEGI) by cultural memory organizations for long term use by the citizens of the USA.

The focus is at-risk government digital information of long term historical significance which is not being adequately preserved via the Web or by other automated means.
Goals

1. Conduct a multimodal scan of at-risk federal digital content, as well as the range of synergistic efforts now being undertaken to preserve it.

2. Develop recommendations for agencies and other government information producers for best practices that can improve the preservability and accessibility of electronic information.

3. Conduct an educational awareness and advocacy outreach program.

4. Surface shared recommendations from a variety of national stakeholder groups for a collaborative agenda to continue improving preservation of electronic government information.
PEGI National Forums Grant

- Funded by $87K grant from the Institute of Museum and Library Services
- Will convene a series of national forums with key stakeholders (especially including fields adjacent to that of librarians, such as subject domain researchers)
- Will identify and surface the elements of a broadly shared national consensus and agenda on specific future steps needed to preserve and provide long term access to electronic government information in the United States.
- Forum activities will include facilitated discussions that will document the relevant knowledge, skills, attitudes, and behavior needed to advance this agenda, and engage disparate stakeholders in planning for aligned activities in the future.
- Beyond the forums held, a key project outcome will be a published report articulating the shared national consensus and agenda on these issues.
Partner Institutions in 2018 PEGI Forums Project

- University of North Texas
- University of North Carolina at Greensboro
- Arizona State University
- University of Missouri
- Stanford University
- Yale University
- Center for Research Libraries
- Educopia

Acknowledgement: Year two (2018) is made possible in part by the Institute of Museum and Library Services grant #LG-88-17-0129-17.
Mini-Forums

- **American Historical Association Annual Meeting** from January 4-7, 2018 in Washington, D.C. - To reach Historians
- **American Association for the Advancement of Science (AAAS) Annual Meeting** from February 15-19, 2018 in Austin, TX. - To reach Scientists
- **Two webinars** will be held during the first six months of 2018.
- **American Library Association Annual Conference** from June 21-26, 2018 in New Orleans, LA. - To reach more Librarians
- **Society of American Archivists (SAA) Annual Meeting** from August 12-18, 2018 in Washington, D.C. - To reach Archivists
- **International Conference on Digital Preservation (iPRES)** from September 24 - 28, 2018 in Boston, MA. - To reach Preservation Researchers
- **IDW2** (joint CODATA, RDA, and WDS) Conference, planned to be held in October 2018 - To reach groups that preserve data
Concluding PEGI National Forum

● Will convene a strongly qualified group of experts and key stakeholders to engage with the issues involved in preservation of electronic government information

● Will be held in mid-December 2018 in Washington, D.C. at CNI.

● Attendees from the various stakeholder communities recruited in the Mini-Forums and Webinar Forums, selected on the basis of both their ability to represent the perspectives of their stakeholder groups and their willingness to actively participate in the National Forum

● Particularly engaged Attendees will be encouraged to assist in writing the draft of the project final report. The draft of the project final report will be made public through the project website and listserv announcements one month in advance of the National Forum

● Final project report will be the blueprint for subsequent mobilization and coordination of synergistic efforts in the future
Part 2:

Group discussion of three questions concerning your thoughts regarding preservation of electronic government information and ways of addressing the issue
Question #1: Are there categories of electronic government information (i.e. digitized records, databases, other digital information types) that you think have long-term value for research purposes? If so, describe briefly in the chat window.
Answers to Question #1 (what should be preserved?)

- Emails, tweets, instant messages, social media postings, audio-visual files
- Any official communications from one part of the government to another
- Dated logs showing who has met with government officials
- Data that serves as the basis for government reports
- Data sets collected by government contractors
- Data on government grants, contracts and sales
- Geodata
- Information obtained through FOIA requests, and the reasons for denials
- Agency marketing ephemera
- Proceedings of government-sponsored conferences
- Private sector interviews with government figures (i.e. C-Span)
Question #2: If you identified any categories of electronic government information in the previous question, do you think that there are any potential risks that might threaten the long-term survival of this information? If so, describe briefly.
Answers to Question #2 (what are the risks for data loss?)

- Loss based on the nature or format of the information product itself
- Loss resulting from change that takes place too rapidly
- Loss resulting from budget/funding problems
- Loss resulting from misunderstanding the needs and requirements for data preservation
- Loss due to political or legal situations
- Loss due to insufficient leadership on the issue
Question #3: Do you think that there are any useful strategies that could lead to better coordination of efforts by academic or public organizations for purposes of digital collection, preservation, and access? Do you see any logical priorities, opportunities, or barriers to such collaborative actions?
Answers to Question #3 (strategies for better coordination?)

- Set up disaster plans for data repositories that lose support or funding
- Educate administrators and political leaders
- Leverage informal networks
- Involve professional networks
- Identify partners such as End of Term archive team, HathiTrust, federal agency librarians
- Vendors could be partners
Part 3:

Invitation to participate in PEGI Project next steps
Consider attending the PEGI National Forum mid-December 2018 in Washington, D.C.

- Join our group of stakeholders if you are interested in helping develop a shared national agenda for the preservation of electronic government information
- Will be held in mid-December 2018 in Washington, D.C., in conjunction with CNI.
- Attendees from the various stakeholder communities recruited in the Mini-Forums and Webinar Forums, selected on the basis of both their ability to represent the perspectives of their stakeholder groups and their willingness to actively participate in the National Forum
Contribute a full PEGI survey

URL for the survey:

Thank you.

Questions?
Contact us:

www.pegiproject.org
info@pegiproject.org
@PEGIProject