TOWARD A SHARED AGENDA:
Report on PEGI Project Activities
for 2017-2019

Authors: Martin Halbert, Katherine Skinner, Roberta Sittel, Deborah Caldwell, Marie Concannon, James R. Jacobs, Shari Laster, and Scott Matheson.

Publisher: Educopia Institute, 235 Peachtree Street, Suite 400, Atlanta, GA 30303

Copyright: 2019


This project was made possible in part by the Institute of Museum and Library Services #LG-88-17-0129-17. The views, findings, conclusions, or recommendations expressed in this white paper do not necessarily represent those of the Institute of Museum and Library Services.

This publication is covered by the following Creative Commons License: Attribution-NonCommercial-NoDerivs 4.0 International. You are free to copy, distribute, and display this work under the following conditions:

**Attribution** – You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner. Specifically, you must state that the work was originally published as Toward a Shared Agenda: Report on Preservation of Electronic Government Information (PEGI Project) Activities for 2017-2019 and you must attribute the copyright holder as Educopia Institute.

**Noncommercial** – You may not use this work for commercial purposes.

**No Derivative Works** – If you remix, transform, or build upon the material, you may not distribute the modified materials.

Any of these conditions can be waived if you get permission from the copyright holder. Your fair use and other rights are in no way affected by the above. The above is a human-readable summary of the full license, which is available at the following URL: https://creativecommons.org/licenses/by-nc-nd/4.0/
“Sloth, like rust, consumes faster than labor wears, while the used key is always bright.”

- Poor Richard’s Almanack
# CONTENTS

## I. CONTENTS

## II. THE PROBLEM STATEMENT

- Introduction to the Project
- Loss of Information in the Digital Age
- Collective Impact Framework

## III. PROJECT ACTIVITIES SUMMARY

- PEGI Project Mini-Forums
- Environmental Scan
- Results of the National Forum

## IV. APPLYING COLLECTIVE IMPACT

- Common Agenda
- Shared Measurement Systems
- Mutually Reinforcing Activities
- Continuous Communication
- Backbone Support Organizations

## V. NEXT STEPS

- What’s Next for the PEGI Project?

## VI. CONCLUSION

## VII. APPENDIX

- PEGI PROJECT: Timeline of Project Activities

## VIII. BIBLIOGRAPHY
THE PROBLEM STATEMENT

INTRODUCTION TO THE PROJECT

In the last thirty years, the rapidly changing landscape of government information creation has shifted from print to primarily digital information and data.\(^1\) In the pre-digital production era, a workflow across hundreds of Federal Depository Library Program (FDLP) libraries accounted for the preservation of and access to significant public government information resources. Federal agencies created content that was disseminated to the Government Publishing Office (GPO\(^2\)), with public access provided through the FDLP. Separately, information identified as archival records was transferred to the National Archives and Records Administration (NARA). In this print-oriented landscape, both agencies bore complementary responsibility for the selection and preservation of government publications and records for long-term access by members of the public.

Today, most government information is produced and disseminated digitally. The distributed system in a digital environment is not nearly as predictable or smooth, and the amount of information has increased

---

\(^1\) Throughout this work, we use ‘government information’ to refer to content or resources produced by or through the direct participation of federal government agencies. At times we refer to specific types of information products, such as reports, publications, or data; or specific categories, such as records or public information. However, in general, we intend to broadly reference these materials.

\(^2\) GPO was originally the Government Printing Office. This paper uses the current agency name, Government Publishing Office, throughout.
exponentially. Digital publishing disrupted the established system for disseminating and preserving government information, which today does not always fit neatly into the traditional “publication” or “record” categories. Individual federal agencies quickly and easily self-publish information directly via the agency’s website as well as outside the .gov web domain on any number of corporate social media platforms — and can just as easily move or remove that same information. If the agency-hosted work does not meet certain official characteristics, no mandate exists for agencies to maintain content themselves, nor to provide it to the GPO or NARA for ongoing care, resulting in information loss. Moreover, a lack of resources can lead to the loss of information that should be maintained under current policy.

At the 2015 Federal Depository Library Council (DLC) Meeting, Dr. Katherine Skinner and Dr. Martin Halbert participated in a panel discussion, “Preserving Federal Information Through Collaborative Models,” during which Dr. Skinner introduced the Collective Impact model. The panel focused largely on preservation of tangible collections with little discussion of born-digital information. The focus on tangible publications ran contrary to Dr. Halbert’s expectations as the majority of government information produced today is born-digital. To better understand the dissemination and preservation of born-digital government information, Dr. Halbert organized two meetings known collectively as the Digital Preservation Summit which were held in conjunction with the 2016 spring and fall membership meetings of the Coalition of Networked Information (CNI).

The first Digital Preservation Summit meeting held in April 2016 engaged national leaders in a structured, facilitated dialogue on at-risk digital government information. It also aimed to explore the development of a shared agenda to address the preservation and access of priority content in this area. The second Digital Preservation Summit meeting, held in December 2016, advanced the goal of organizing the work more formally. Of those in attendance, six committed to continued efforts; these six became the first members of the Preservation of Electronic Government Information (PEGI) project steering committee.

---

7 A full report is available at: https://digital.library.unt.edu/ark:/67531/metadc826639/ [https://perma.cc/42ZZ-QW2M].
8 The initial PEGI Project steering committee was comprised of Martin Halbert, Roberta Sittel, Marie Concannon, James R. Jacobs, Shari Laster, and Scott Matheson. Lynda Kellam served on the steering committee for most of 2018, and Deborah Caldwell joined the steering committee as the project assistant in 2017.
In May 2017, several members of the newly formed PEGI group attended the Libraries+ Network™ meeting hosted by the Data Refuge originators, The Mozilla Foundation, and the Association of Research Libraries (ARL). This meeting continued to address themes surfaced in the Digital Preservation Summit meetings, reinforcing the need for continued discussions around preserving born-digital government information. At the culmination of this meeting, Dr. Halbert, then with the University of North Texas, offered to take the lead on an environmental scan and pursuit of grant funding to address collective action.

Later that month, the University of North Texas received a National Leadership Grant from the Institute of Library and Museum Services (IMLS) to conduct a series of forums to engage librarians, archivists, researchers, and other stakeholders in conversations about the importance of preserving electronic government information for long-term access. The IMLS grant supported activities culminating in the December 2018 PEGI Project National Forum, held in conjunction with the CNI fall membership meeting, which aimed to:

- strengthen relationships among key stakeholders engaged in the preservation of electronic government information
- jointly develop a shared understanding of preservation activities currently underway
- explore shared goals and identify potential shared measurements of progress toward those goals
- identify and prioritize activities specific stakeholders might engage in to lower barriers
- test a possible “shared agenda” approach for ongoing cross-stakeholder collaboration

**LOSS OF INFORMATION IN THE DIGITAL AGE**

In the print era, access to and preservation of government information was synonymous with access to and preservation of print copies of documents. Bibliographic control, while important for access, could be gained over time; and print is an inherently stable format. In the digital era, access and preservation have become more complex with the possibility of accessing the same electronic information from innumerable locations via the internet. Preservation of digital content is possible through ongoing replication of encoded information over time, but collection and control of content is a critical prerequisite for any steps toward preservation.

Electronic information is also vulnerable to additional threats to access and preservation. Unlike physical artifacts such as bound print volumes, which can survive for decades with minimal intervention, digital
information can all-too-easily wink out of existence in the absence of a preservation plan, workflow, and infrastructure. Digital information can also degrade in more subtle ways. For example, a website redesign can orphan content by no longer linking directly to pages, even if the content is still hosted as it was before. Electronic content can be accidentally or intentionally deleted or altered far more easily, especially if only a single instance of the content exists.¹¹

But information loss is not only a technological issue. There are a number of documented instances in which significant amounts of electronic information have disappeared from government websites and repositories, and are therefore no longer easily accessible to the public. Anne Heanue, at the Washington Office of the American Library Association (ALA), documented information loss from 1981-1998 in a chronology called Less Access to Less Information By and About the U.S. Government.¹² Starting with the 1980 Paperwork Reduction Act (PL 96-511) passed by the Reagan Administration, Less Access contained examples of information loss as well as government efforts to privatize information and restrict access via policy or obfuscation, by malice or accident. Heanue reported that between 1982 and 1987, one in every four federal government publications was eliminated altogether, and thousands of others that were once distributed for free became available only for sale. The blog Free Government Information (FGI)¹³ has digitized Less Access and has been keeping a running tally of information loss in honor of Heanue’s work since 2005.

In some of the following examples, “information loss” or “data loss” refers to restricted or eliminated public access, rather than the content itself being irretrievably lost or purged. The overall effect is to lessen public access to government information, even in cases for which some of the underlying data can eventually be uncovered.

- In early 2013, NASA abruptly took its Technical Report Server¹⁴ (NTRS) offline due to an off-hand comment from Representative Frank Wolf over concerns that export-controlled information was contained in the collection. The database was partly back online six months later, but NTRS

---

¹² Less Access to Less Information By and About the U.S. Government, [https://freegovinfo.info/less_access](https://freegovinfo.info/less_access).  
¹³ The blog Free Government Information (FGI) is co-authored by James R. Jacobs, a member of the PEGI Project steering committee.  

- In the President’s FY 2012 Budget Request,\footnote{Office of Management and Budget, Fiscal Year 2012, Terminations, Reductions, and Savings, Budget of the U.S. Government, https://www.gpo.gov/fdsys/pkg/BUDGET-2012-TRS/pdf/BUDGET-2012-TRS.pdf [https://perma.cc/2EUQ-F2JM].} the National Biological Information Infrastructure (NBII) program was terminated and all resources, databases, tools, and applications within the site were removed. NBII coordinated a broad array of standardized information at the federal level about biodiversity and ecosystems. Several libraries attempted to collect and preserve the website,\footnote{National Biological Information Infrastructure, http://wayback.archive-it.org/2361/201205233133/http://www.nbii.gov/portal/server.pt?open=512&objID=236&mode=2&cached=true} but much of the dynamic functionality and information was lost.

- In July 2018, the National Guideline Clearinghouse and the National Quality Measures Clearinghouse from the Agency for Healthcare Research and Quality (AHRQ) were taken offline due to a loss in funding. Access is currently not available\footnote{James A. Jacobs, “Preserving What’s Gone--The Healthcare Guidelines Case,” Free Government Information, 2018, https://freegovinfo.info/node/12925 [https://perma.cc/BH85-SQJR].} to these critical resources for doctors and others in the medical community.

- Over the past two years, Environmental Data Governance Initiative (EDGI)\footnote{Environmental Data & Government Initiative, https://envirodatagov.org/ [https://perma.cc/P72Z-78WR].} has been monitoring federal websites\footnote{Environmental Data & Government Initiative, https://envirodatagov.org/website-monitoring/ [https://perma.cc/V79V-75TT].} and analyzing how the information has changed over time. EDGI found that several agencies have significantly changed or deleted websites during this period. For example, the Environmental Protection Agency (EPA) discontinued updates to its climate change website; removed its “International Priorities” page; removed Climate Action Plans and Links from the National Park Service’s Climate Friendly Parks Program Website; and removed “Climate and Energy Resources for State, Local, and Tribal Governments” website, later replacing it with “Energy Resources for State, Local, and Tribal Governments.”\footnote{These are but a few examples of information and data loss. For a more in-depth look, see McDermott, Patrice. Who Needs to Know? The State of Public Access to Federal Government Information (2007) (http://www.worldcat.org/oclc/310969160).}

**COLLECTIVE IMPACT FRAMEWORK**

As the PEGI Project began its formal efforts in 2017, we quickly encountered resonant examples of
concern around the potential loss of electronic government information. It became apparent to us that activities to preserve electronic government information benefit from multi-institutional collaborations. Collaborative efforts such as EDGI and the End of Term (EOT)\textsuperscript{22} web archiving project focus on specific types or time frames of government information and are effective in their preservation efforts because of coordinated efforts and a common goal. In recognizing the strength of coordinated efforts, the PEGI Project adopted Collective Impact as the conceptual model for exploring collaborative possibilities within this problem space.

For the project framework, we chose to explore the potential applicability of Collective Impact because of its success in mobilizing disparate groups around a targeted effort. In their 2011 article, Kania & Kramer write, “complex problems can only be solved by cross-sector coalitions.”\textsuperscript{23} Following this approach, the PEGI Project engaged diverse groups in conversation to explore the development of a shared agenda and build a cross-sector coalition. Later in this paper we discuss Collective Impact in depth and describe our assessment of its application to the preservation of born-digital government information.

---

\textsuperscript{22} End of Term Web Archive, http://eotarchive.cdlib.org/ [https://perma.cc/2NL8-E94A].

PROJECT ACTIVITIES SUMMARY

PEGI PROJECT MINI-FORUMS

Beginning in Fall 2017 and continuing through Fall 2018, the PEGI Project hosted a series of mini-forums and events intended to raise awareness of the need for active preservation of born-digital government information, elicit feedback on the scale and scope of the problem, and lay the initial groundwork for an approach to possible solutions.

In several of these mini-forums, we used a set of three questions to provoke consideration and discussion of this topic. Responses were also elicited from participants following the mini-forums, and from other interested parties. The questions posed were:

1. What categories of electronic or born-digital government information have long-term value for research purposes?
2. What risks might threaten the long-term survival of this information?
3. What strategies could lead to better coordination of efforts by academic or public organizations for purposes of digital collection, preservation, and access; and what are some logical priorities, opportunities, or barriers to such collaborative actions?

While these questions shifted depending on the audience, the intent remained consistent throughout: to encourage thinking about the context for long-term access.

In general, we found a divide between responses from researchers and academic practitioners, and responses from library and information professionals, archivists, and others active in these fields.
Academic researchers tended to produce fairly detailed responses to the first question, particularly when prompted to consider digital parallels to research using tangible sources. In particular, those with expertise in history and the humanities often mentioned government sources that are not public information dissemination products, such as emails, dated meeting logs, and notes or memos summarizing meetings.

In contrast, and in keeping with fairly widespread notions of how libraries and archives collect and organize materials, information experts tended to enumerate broad categories. Examples included: data-gathering protocols and policies, standards or specifications, and metadata; gray literature such as presentations and contractor-funded reports; records of administrative rulemaking including public comments on proposed regulations; and court opinions, briefs, and other records related to legal proceedings.

Regarding risks and threats, researchers tended to focus on perceived threats from political forces, though some also acknowledged interoperability, technical obsolescence, and metadata as additional factors. Librarians, archivists, and other information professionals more frequently identified funding and infrastructure as primary challenges. Some additionally pointed out shortcomings in agency compliance with records management policy or the inadequacies of these policies. Those responsible for information creation may lack resources or encounter difficulty in developing workable and scalable solutions for content collection and description.

For the third question, information practitioners’ responses varied widely, but tended to focus on two primary categories: improvements to information policy, including funding and a national mandate for collecting, organizing, and preserving content, supported with improved guidance and support for information creators; and the mobilization of emerging network-based approaches to collecting, curating, and preserving born-digital government information and data on behalf of the public interest. Yet our conversations at mini-forums were unsuccessful in drawing extensive insight from the research community. We collected some feedback of a general nature, but PEGI’s continuing work will need to use other methods to draw further comments to this very important question.

What follows is a summary of PEGI Project mini-forums held during this period. Throughout the year, the PEGI Project also presented posters and panel discussions at other professional conferences and events, enumerated in the Appendix to this report.

**Federal Depository Library Conference (FDLC), October 2017**

During the lead-up to the mini-forum series, the PEGI Project facilitated an active discussion with attendees of the 2017 Federal Depository Library Conference and Depository Library Council Meeting. Government information librarians and other professionals discussed and responded to a series of questions about born-digital government information access and preservation, reflecting
on known and emerging risks and possible strategies for mitigation.

These discussions were later summarized in a series of PEGI Project blog posts.  

American Historical Association (AHA) Annual Meeting, January 2018

At the first PEGI Project mini-forum, a small group of historians discussed long-term research needs for born-digital government information. Attendees included several students and practicing historians, as well as government employees from the local, state, and federal levels. The participants in this group primarily focused on the archival record of government decision-making. For example, several pointed out the necessity of information gathered in the process of decision-making, such as emails and other internal reports. Several participants noted that certain time periods were especially at risk for loss of historical materials, such as the mid-1990s into the mid-2000s.

American Library Association (ALA) Midwinter Meeting, February 2018

At ALA Midwinter, the PEGI Project facilitated an active discussion with forum attendees, including several dozen members of ALA’s Government Documents Round Table (GODORT). Much of this discussion reflected similar concerns and insights as the FDLC mini-forum.

American Association for the Advancement of Science (AAAS) Annual Meeting, February 2018

The PEGI Project mini-forum included outreach to active researchers and practitioners who rely on federally-created and federally-funded data. We were able to engage with individual attendees between sessions to garner interest. Discussions centered on research data management, including government-produced data, and informed our understanding of how researchers in the sciences understand the sources and vulnerabilities of their data sources.

PEGI Project Virtual Mini-Forum, March 2018

We organized a virtual mini-forum to allow interested participants to consider and reply to the three questions. Attendees included librarians, archivists, students, and government employees, with a wide range of disciplinary expertise. Important opportunities for coordinated effort emerged during group discussion, and brought a number of projects to our attention as possible models for

cooperative activity.

**Organization of American Historians (OAH) Annual Conference, April 2018**

The PEGI Project’s participation at the OAH Annual Conference focused on tabling in the exhibit hall to raise awareness of the need for a coordinated strategy for collecting and preserving born-digital government information. We spoke with independent scholars, vendors, and other attendees about the current landscape for this work, and came away with more insight on researcher perspectives.

Due to budgetary reasons and timing of activities, this event substituted for participation in the Research Data Alliance (RDA) and International Data Week (IDW) co-conference in Gaborone, Botswana.

**PEGI Project Webinar, May 2018**

For the second virtual mini-forum, we invited key activists and stakeholders to reflect on the PEGI Project “three questions.” Presenters included:

- Rachel Mattson, Curator of the Tretter Collection for GLBT Studies at the University of Minnesota Libraries & founder/co-leader of the Digital Library Federation's interest group on Government Records Transparency & Accountability
- Brandon Locke, then Director of LEADR at Michigan State University & founder & co-organizer of Endangered Data Week
- Bernard F. Reilly, then President of the Center for Research Libraries
- Justin Schell, Director of Shapiro Design Lab & member of the Environmental Data & Governance Initiative (EDGI)
- Bethany Wiggin, Founding Director of the Penn Program in Environmental Humanities
- Patricia Kim, then Program Coordinator of the Penn Program in Environmental Humanities
- Heather Joseph, Executive Director, SPARC

These speakers discussed highly relevant approaches and considerations for motivating and organizing efforts to preserve government information. The long-term cost of inaction is certainly one motivator, but both the scale of the challenge and the invisible nature of the infrastructure needed to address it can make it hard to know how to proceed with any given solution. Seeking out collective investment in data at a more personal, human-navigable scale can help to push through some inertia and create avenues of engagement. Building coalitions with projects leading change in related areas of concern, such as Open Access, can amplify progress that makes our own efforts
more likely to succeed.

We highly recommend viewing the recording of this informative discussion.\textsuperscript{25}

**ALA Annual Conference, June 2018**

Because the ALA Midwinter Meeting mini-forum in February 2018 was co-hosted by GODORT and focused on government information librarians and professionals, our mini-forum at the Annual Conference was targeted to reach conference attendees from other communities of practice, including research data management and scholarly communications. This meeting, which largely echoed prior discussions with the government information librarianship community as summarized above, drew librarians from outside the United States, representatives of national libraries, and other professionals who do not necessarily specialize in government information but recognize the relevance of this issue for their own areas.

**American Association of Law Libraries (AALL) Annual Meeting, July 2018**

At AALL, we conducted a mini-forum to raise awareness and gather input from the law library community. The well-attended session led to a lively discussion of past efforts and known access issues, and allowed us to gather additional perspective about legal information and the needs of law librarians, lawyers, and judges.

**ARCHIVES * RECORDS 2018, August 2018**

At the joint meeting of the Society of American Archivists (SAA), the Council of State Archivists (CoSA) and the National Association of Government Archives and Records Administrators (NAGARA), the mini-forum featured lively, small-group discussions with archivists and records managers, many of whom had extensive experience with records management for state and local governments. Issues raised in this discussion had a significant impact on how we framed discussions engaging with librarians and archivists going forward.

**iPRES, September 2018**

Project participation at the International Conference on Digital Preservation (iPRES) included a panel discussion focusing on how policies affect access to and preservation of information. The panel, which included David Walls and Cindy Etkin from GPO, Carla Graebner from Simon Fraser University, Frederick Zarndt, who is a preservation consultant, and Edward McCain with the Reynolds Journalism Institute, also discussed collaborative approaches to addressing policies that are both positive and negative for the preservation of electronic government information. Speakers from GPO explained how policies guide programs and the importance of engagement with policy-setting.

Speakers with international experience highlighted policies that have helped and hindered electronic information preservation efforts. This session reinforced the importance of raising awareness about the need for good policy.

**ENVIRONMENTAL SCAN**

In Spring 2018, the PEGI Project commissioned a scan on behalf of its participating institutions to document digital preservation activities currently taking place and preservation gaps in the government information ecosystem. Conducted by Sarah Lippincott with guidance from members of the PEGI Project steering committee and additional support provided by Educopia, the goals of the scan included defining the information realm, surveying communities of interest, and preparing a report detailing the organizations, services, and infrastructures currently in place to preserve government information. This scan focused on born-digital data and information produced by the federal government of the United States.

The resulting report describes the landscape of initiatives within and outside of the federal government that aim to disseminate and preserve government information and data. It first describes government-led initiatives, from dissemination through official agency websites to publication on third-party platforms. Next, it considers the range of initiatives that have emerged in recent years outside of government to address perceived gaps and vulnerabilities in the federal government’s curation initiatives and to add value to publicly available information and datasets. It briefly touches upon initiatives that focus on advocacy, awareness, or education, rather than on directly providing preservation and access. The report goes on to address the policies and infrastructures undergirding both government-led and non-government initiatives. It concludes with a brief summary of gaps and recommendations for collective action. Each section contains representative examples but does not contain an exhaustive list of initiatives relevant to federal government information.

In brief, the Environmental Scan project affirmed that much of the government information that is made publicly available is distributed directly by the agencies that create it via their top-level websites and subdomains. However, this is far from the end of the story, as information and data may also be hosted or replicated through a variety of other government-administered platforms, from data portals that aggregate content from multiple agencies to social media platforms that broadly disseminate government-created information. These government-administered sources include entities such as:

- single-agency websites

---

26 The summary that follows is substantially condensed from the full Environmental Scan, written by Sarah Lippincott in October 2018, and published by Educopia in December 2018. We are grateful to Sarah for her detailed research and keen insight into the needs of this effort. See: https://educopia.org/wp-content/uploads/2018/12/PEGIEnvironmentalScan-1.pdf [https://perma.cc/X5XG-ZFFD]
The proliferation of born-digital government information requires new approaches to collection and preservation. The web has made access to government information easier than ever, as digital publishing becomes easy and instantaneous. However, the sheer volume of information being produced, and its decentralized distribution through hundreds of government domains, has disrupted traditional models of preservation.

To make sense of the deluge, dozens of non-governmental projects have emerged to curate and preserve born-digital government information. These projects have developed in response to a variety of needs and concerns and have therefore adopted heterogeneous goals and approaches. In general, they function independently of the government agencies whose data and information they archive.

It is important to note that, though information collection and preservation efforts cover a wide range of government agencies, none is comprehensive. At their most ambitious, they aspire to duplicate and provide access to the most at-risk information, as identified by qualified experts in the field. Other projects have more modest or specific goals, such as efforts to preserve and disseminate Census Bureau data from a particular year.

The breadth of agency content represented in non-government-administered repositories is impressive. This scan identified non-government repositories that preserve information and data from at least 58 different government agencies and sub-agencies and the executive branch, including all 13 Federal Statistical System (FSS) agencies. Their approaches to collection and curation are informed by a variety of goals and desired outcomes, from aggregating information relevant to specific research communities to foster greater government transparency.

The challenges of digital preservation are not unique to government information. Cultural heritage organizations make difficult decisions about what born-digital materials get preserved, given technology and resource constraints; archivists and librarians must contend with the difficulty of discovering information in distributed repositories, and curators face limitations of current web archiving tools, which cannot easily collect files hosted in dynamically queried databases, unless the databases include site maps or structured URLs.

Preserving government information is a long-term responsibility that requires ongoing coordination between technologists, librarians, archivists, and disciplinary experts, both within and outside the
government, with substantial input and advice from public interest organizations and other communities. Collaboration between stakeholders is essential to ensure that information is preserved responsibly, safeguarded against technological or other failures, and broadly discoverable and accessible. Given the immense volume of digital government information, developing preservation priorities requires consultation between digital archivists and disciplinary experts who can identify the most important information sources for their fields. Collaboration with technologists ensures that preservationists can adopt the most robust methods for curating new media content. And a holistic community effort is necessary to ensure that the failure or sunsetting of individual projects does not lead to catastrophic losses.

Despite the many accomplishments of these initiatives, the greater issue of preserving and providing long-term access to government information may be best addressed through a series of coordinated efforts. Going forward, several pressing needs stand out. There is a clear need for an organizing body to leverage the existing distributed infrastructure and expertise into a coordinated effort by setting priorities, staying aware of imminent threats, parceling out responsibility, and advocating for greater government information transparency. There is a clear need to reevaluate government policies, regulations, and legislation regarding the deposit of digital information products with federal publishing and record-keeping agencies. There is a clear need for the implementation of open standards and interoperable metadata that empower external organizations to analyze the full range of government information. Government agencies should be held responsible for producing and disseminating their information products in standards-compliant formats and repositories.

The challenges are great, but the risks of failing to address them are profound. The public stands to lose direct or unmediated access to decades worth of historically and scientifically significant information. The actions that the federal government, and the non-governmental organizations addressing these issues, take will have consequences for future generations of scholars, scientists, and the public at large.

**Results of the National Forum**

The PEGI National Forum, held December 9-10, 2018 in Washington DC, was the capstone to our series of mini-forums. Whereas mini-forums were planned as one or two hour-long sessions for the purpose of gathering responses to our three questions as posed to specific communities, the National Forum was a two-day event designed to bring together diverse groups to take a deeper dive into the issues. National Forum participants were selected based on their track record of leadership, experience, or interest in electronic government information preservation, as well as for their ability to effectively represent their

---

27 A detailed summary of the National Forum process and outcomes can be found in our report, available at: [https://educopia.org/pegi-project-national-forum-summary-and-report/](https://educopia.org/pegi-project-national-forum-summary-and-report/). We are grateful to all forum participants and acknowledge that their remarks and insights have been essential in informing our understanding of the landscape for this work. To encourage the free flow of conversation, participants at the National Forum agreed to proceed under the Chatham House Rule; therefore, we do not attribute statements made by participants.
respective fields and industries. This forum brought together key stakeholders from federal agencies, state government leaders, open government advocates, archivists, librarians, policy analysts, data specialists, scientists, journalists, digital preservationists, historians, and others.

To explore the feasibility of applying the Collective Impact model as a framework for collaborative action, Dr. Katherine Skinner of the Educopia Institute structured the National Forum as a progression of presentations and discussions, brainstorming activities and small group work, culminating with a creative exercise to plan projects that may be suitable for implementation through cross-institutional collaboration.

To set the stage for this dynamic event, participant groups were asked to select a metaphor to represent how they envision the idea of preserving electronic government information. One metaphor came from a group that chose an image of a lone, bright key amidst a pile of old, rusty keys. They quoted a proverb from *Poor Richard's Almanack*: “Sloth, like rust, consumes faster than labor wears, while the used key is always bright.” This illustrates that neglect of digital files can contribute to their demise, whereas digital objects that are well-used have a better chance of lasting usefulness. This contrasts with principles for the preservation of paper, which typically discourage excessive handling.

With participants representing a wide array of institutions and professions, it was important to ensure that communication at this forum be as clear and effective as possible. While it would not have been possible to achieve complete agreement on the meaning of every relevant term such as ‘information,’ ‘metadata,’ or ‘machine-readable,’ an exercise to identify ‘trouble terms’ was useful in demonstrating how extensive the differing perspectives could be.

For example, words such as ‘access’ and ‘preservation’ have multiple commonly used senses aside from the technical definitions. Librarians and archivists have different but equally valid professional definitions of what constitutes access. In the same way, the term ‘preservation’ could indicate any point on a broad spectrum of end goals and involve vastly different underlying processes and requirements. Even the ‘government’ part of ‘government information’ is not understood uniformly. Some participants stressed that the universe of government information should include resources created by private contractors hired by agencies and paid with public funds, and possibly some federal grant recipients. However, various factors affect whether government information produced on contract by private entities legally falls into the public domain.28

After exploring differences in language and terminology, the group was next asked to look for areas of agreement and form some common goals. Indeed, one of the most important outcomes of the National Forum was to build consensus on shared priorities for collective action. During this exercise, many suggestions for goals surfaced. All were then collected and rearranged to reveal patterns. The result was the identification of six directions that seem ripe for action:

28 For an overview of factors affecting the copyright status of works created under a federal contract or grant, see: https://www.cendi.gov/publications/FAQ_Copyright_30jan18.html [https://perma.cc/HJJ9-X2KD]
1. Build public awareness and coordinate advocacy campaigns for the preservation of electronic government information.
2. Improve methods for identifying and prioritizing information for preservation.
3. Seek solutions that closely align with public needs, making sure they are both diverse and inclusive. Provide access opportunities that account for inequities that result from the digital divide.
4. Improve coordination among existing government information preservation projects and avoid duplication of effort.
5. Prioritize projects that are financially sustainable for the long term.
6. Work in the public policy arena to promote policies that improve access to and preservation of government information, and improve guidance and support for policies that are already in place.

After agreeing upon these six directions, participants were asked to identify shared metrics that could track progress. Across many of the groups, a common stumbling block emerged. Metrics are difficult when the universe of what is being measured is undefined or amorphous. In other words, it is difficult to measure how close we are to a goal – presumably, preserving all government information – if it is not possible to state how much government information even exists. Taking a more qualitative approach, it was noted that subjective goals such as “a stronger and more sustainable democracy” cannot be easily measured.

What follows is a synthesis of the five most significant types of shared metrics proposed during this National Forum.

**What is Preserved?**

This metric focuses on discrete counts of digital materials. Participants listed ways to count the amount of born-digital information that is preserved or made available.

**Usage**

This metric involves measuring the use of digital government information. Tracking anonymized usage could help by creating a surrogate for the value of these resources, potentially increasing our ability to identify what gets preserved and what requires additional attention.

**Systems and Infrastructure for Preservation**

This metric involves counting the systems and infrastructure in place for preservation activities. For example, numbers might be gleaned from a registry of official mirrors within and outside institutions. There are numerous ways to track archived content stored at host institutions.
Alternately, we might count agencies that partner with Data.gov for registering descriptive metadata for their data products. It may also be informative to count non-governmental organizations that partner with agencies on preservation activities.

**Policy**

Public policy on government information preservation can include regulations directed toward agencies as information creators. It also involves regulations directed toward institutions which are mandated to oversee government information, provide access to it, and preserve it for the long term. Public policy is also significant in determining funding for preservation activities.

**Stories that Demonstrate Value and Impact**

News stories and ongoing press coverage about the value and positive impact of access to government information can increase public support. This kind of advocacy can help collaborative preservation projects by attracting funding and other support necessary for long term project sustainability.

In the discussion wrap-up, participants generally agreed that a primary challenge for framing shared metrics is that preservation activities have complex contexts and many dimensions for determining what constitutes progress.

For the final portion of the forum, each table group was asked to propose a project that would address preservation issues raised in earlier discussions. Projects were pitched to the full room, complete with a title, organizational structure, potential partners and a description of activities leading to one or more goals, keeping in mind that it would be necessary to measure progress toward goals with shared metrics. The purpose of this activity was not necessarily to create actionable blueprints or to commit organizations to become partners, but rather to demonstrate the many opportunities for creative action that are available to us at this juncture. Within the short time that groups had to complete this exercise, plans were drafted that have the potential to inform some of our post-grant work.

---

29 Data.gov, managed and hosted by the U.S. General Services Administration, Technology Transformation Service, is a data catalog meant to point to various datasets created and hosted by federal agencies. See [https://data.gov](https://data.gov) [https://perma.cc/FE7K-9UUY].
APPLYING COLLECTIVE IMPACT

The original 2011 article by Kania & Kramer that proposed the Collective Impact model identified “five conditions that together produce true alignment and lead to powerful results: a common agenda, shared measurement systems, mutually reinforcing activities, continuous communication, and backbone support organizations.” In planning for the IMLS grant, we decided to explore the potential applicability of Collective Impact as a way to frame project activities and inform possible outcomes.

While subsequent research on the effectiveness of the Collective Impact model has produced literature that challenges and amends Collective Impact, our project engaged with the Collective Impact model as it was originally proposed. As described above, the PEGI Project mini-forums allowed the project team to involve a variety of stakeholders in discussions to solicit feedback about specific concerns of loss of historically significant electronic government information. The National Forum featured structured and facilitated activities designed with the Collective Impact framework in mind. Expectations for the National Forum were informed by the mini-forums as well as the PEGI Project team’s understanding of Collective Impact.

During preparation for the National Forum, some members of the PEGI Project team were confident that we could satisfy the five conditions for Collective Impact as we worked through the various activities. Other team members had a more cautious outlook and saw the National Forum as an opportunity to build relationships with those in attendance, but not necessarily to find common ground on all five conditions.

In summarizing and digesting the results of the National Forum, we concurred that although it was at times challenging to achieve consensus, we did identify distinct areas of agreement through the Forum’s carefully-planned discussions and activities. Yet the implications for future organized group activities
and projects were not entirely clear, particularly when we tried applying the framework of Collective Impact to determine our next steps. In the reflective period following the National Forum, we began to examine our assumptions regarding Collective Impact as initially framed in 2011 as we further explored critiques, successes, and adaptations of Collective Impact.

What follows, then, is a brief assessment of our current understanding of the potential for Collective Impact efforts within the born-digital government information preservation community as we engaged with it through the course of project activities, most significantly as part of the National Forum. This assessment is informed by recent literature on Collective Impact, most significantly Cabaj & Weaver (2016) who propose revisions to the five conditions for Collective Impact:

- Common Agenda into Community Aspiration
- Shared Measurement into Strategic Learning
- Mutually Reinforcing Activities into High Leverage Activities
- Continuous Communication into Inclusive Community Engagement
- Backbone Organizations into Containers for Change

We have followed Kania & Kramer’s framing in this assessment of our initial project outcomes; however, we intend to continue considering the usefulness of centering our approach on Cabaj & Weaver or other re-framings of Collective Impact, going forward.

**COMMON AGENDA**

The stakeholder forums conducted by the PEGI Project demonstrated that there is broadly shared concern in many groups that historically important electronic government information in the United States is currently at risk of loss, and in some cases has already been lost. Close consultation with these stakeholder groups also reveals that many groups have different understandings of the problem and what the ultimate goal and ideal mechanism should be for preserving electronic government information. The Collective Impact model acknowledges that such differences will inevitably exist, and will often persist over time. The model indicates that such differences can be surfaced and reconciled with a set of primary goals for which all stakeholders can agree that progress, as measured by the agreed upon metrics, is desirable and can be achieved.

The activities of the National Forum were designed to move participants toward a common agenda by providing the opportunity to explore shared concerns of information loss. This exploration resulted in common themes developed among attendees. The most consistent theme, which also happens to be one that serves as a major hurdle in defining success for preservation efforts, is the difficulty of knowing

---

the universe of electronic government information being produced and disseminated, and thus of knowing what’s truly at risk. Other common themes include concerns around records retention policies and applied practices; risks due to format limitations or inadequate metadata; political risk; lack of truly public information; and the idea that trusted digital repositories mitigate risk. These themes are reflective of work being done by groups represented at the National Forum and offer reasoning that continued conversations around current activities and perceived risk could produce a common agenda.

The difficulty we experienced in seeking an articulation of common grounds for shared aspirations with respect to measurable progress is emblematic of the complexity of the problem space. In reviewing the common aspirations surfaced through discussion (described under “Results,” above), all six of the agreed-upon directions pertain to the space in which the problem is solved, and not the nature of the problem itself. A solution statement such as “collect and preserve all born-digital government information” is insufficiently specific; however, more specific instantiations of this proposal would fail to generate buy-in by all stakeholders.31

With the breadth of communities involved in this work, we see the possibility to reframe a common agenda as building shared aspirations for change, following Cabaj & Weaver (2016). We can work toward improved approaches to triaging and prioritizing information that needs to be collected and preserved; raise the profile of issues such as equitable access and sustainability; and achieve more responsive coordination of existing efforts, whether these efforts focus on collecting born-digital content or raising awareness of the need for more responsive policy.

We do think it is possible that a more formal agenda-building effort may succeed within a subcommunity. For example, many institutions are engaged in web harvesting efforts. With the development of Cobweb32 and other tools intended to leverage the impact of web harvesting at scale, a concerted effort among invested organizations could lead to more effective collection, description, and preservation practices.

31 For example, some participants in the National Forum expressed deep concern about issues surrounding transparency and loss of access to information, and indeed these issues were surfaced repeatedly in mini-forums as well as the Environmental Scan. However, for some of our agency partners, a greater concern might be resources for effective compliance with information policy or records management requirements; simply meeting these standards can be a challenge without sufficient resourcing, let alone exceeding them to improve public access. Therefore, the shared common ground is in seeking effective policy solutions, not directly in increasing the amount of government information and data available to the public.

32 Cobweb is a platform currently under development by the California Digital Library. This tool is intended to facilitate coordination of web harvesting efforts. See: https://www.cdlib.org/services/collections/webarchiving/cobweb/ [https://perma.cc/Z9LJ-9K9A].
**Shared Measurement Systems**

Shared measurement is the Collective Impact condition that has caused the most discussion and concern throughout the activities of the PEGI Project to date. The unknown scope to the problem and the wide variety of current preservation activities both exacerbate the ability to define and apply clear metrics to the work of preserving electronic government information. One desired outcome of the National Forum was to identify shared measurement systems that clearly illustrate the success of preservation activities.

A number of metrics could potentially gauge technical progress toward preservation of electronic government information, although all of these metrics require additional work to reach more standardized definitions. For example, tracking total bytes preserved, or number of URLs or files preserved, requires rigorous agreement on what to count and how to count it, such as what constitutes adequate preservation and description, and what types of files to include in the tally. However, such an agreement could eventually be achieved.

As discussed above, however, the breadth of metrics with which the invested communities are concerned is significantly larger than technical progress. Ways to count preserved content and its usage were surfaced, as were ways to quantify infrastructure development, policy change, and public awareness and investment in these issues. It might be more meaningful and manageable to look at measurements for individual projects, such as End of Term (EOT), rather than trying to apply a single, standard measurement. Incremental success is still success.

If we were to only engage with efforts invested in the measurement of technical progress, we could lose the reinforcing impact of stakeholders invested in associated efforts with a less quantifiable but nevertheless socially significant impact in this problem space. Seeking approaches for groups of stakeholders to strategically evaluate progress could increase the overall impact of efforts, while not precluding progress within a subcommunity toward agreement on how to quantify preservation.

**Mutually Reinforcing Activities**

If stakeholders can develop a common agenda in the form of a generally agreed-upon future desired state of affairs, the Collective Impact model shows how to synergize activities between disparate but aligned groups. To provide an oversimplified example, if the only thing we collectively care about is bytes preserved, a steadily increasing collective tally would suffice to measure progress toward the shared goal. However, government information and data preservation activities take place in a complex landscape, with many potential ways to understand success. Moreover, simply tallying bytes does not address shared priorities of transformative changes in the availability of born-digital government information.

Here we see the possibility of a more nuanced approach to mutual reinforcement. One consistent finding that has emerged in our conversations with those working in this space is the need for some
organizations to do what others cannot. An engaged, heterogenous, and diverse community of stakeholders benefits from understanding each others’ efforts and seeking insight into how each project can increase its own impact, in light of what others are doing.

**Continuous Communication**

Another condition of the Collective Impact model is that the relevant set of actors must regularly come together to communicate and collaborate over an extended period of time to build shared trust. Potential partners need time to see that their own programmatic interests are respected, and that decisions are made using rationales that they support in ways that are not biased toward other groups.

In their work Kania & Kramer highlight difficulties that disparate groups encounter in reaching a shared understanding of vocabulary in order to discuss the problem issues, an experience we initially shared as we began to discuss preserving electronic government information with various stakeholder groups through the mini-forums and other events. However, we found that the National Forum participants readily identified barriers and discontinuities in vocabulary and terminology and were able to outline characteristics of potentially shared definitions, or, in some cases, agree that particular terminology was simply too fraught and needed to be worked around entirely. Absent a strongly-defined common agenda, continuous communication may be less contentious, but it is also potentially less productive.

Throughout PEGI Project activities, conversations have been key. In the mini-forums, our webinars, the National Forum, and at other venues, the team engaged with diverse stakeholders. This engagement has provided us with windows into how different groups perceive the challenges of preserving electronic government information, where they agree, and how different experts assess the work already taking place. These conversations have allowed relationships to form, and provided space for connecting potential collaborators. Both before and after the National Forum, the team has been called upon to share ideas, serve as potential collaborators, and offer advice. The PEGI Project has gained recognition as a group aware of the work being done to preserve electronic government information, and as a partner in helping others to find their way in this space.

We have repeatedly seen the productive value of consensus-building through continuous communication between potential collaborators, which suggests the need for further convenings focused on born-digital government information and data. These convenings should be structured in ways that bridge boundaries and build shared understanding and consensus about the issues, while deeply engaging with the expertise manifested within these communities of practice.

---

33 For example, many experts who are highly invested in long-term access to born-digital government information do not have the resources or administrative support in their own organizations to build collections; others who are working in this space may be operating within institutions that limit their ability to seek public or funder support for additional efforts.
BACKBONE SUPPORT ORGANIZATIONS

An essential component of the Collective Impact model is the presence of an organization dedicated to convening and facilitating collaboration among the groups participating in an initiative. Along with building a common agenda, establishing or agreeing upon a backbone organization is frequently a priority for initiatives that have achieved success through Collective Impact practices.

While adequate funding remains the most significant challenge for preserving electronic government information, we see other substantial challenges in the formation of a robust backbone support organization. For one, absent sufficient agreement on a common agenda, keeping independent projects on track toward a common goal is not practical. Consensus on the need for raising awareness is clear; however, the issues defined in the National Forum relate to the broad sphere of information policy, not just the need for additional efforts that result in tangible metrics of progress with preservation. Moreover, the relative scarcity of non-federal investment into preservation of government information suggests there is not yet sufficient support for providing resources for a dedicated backbone organization.

We agree that success within this arena can be more dispersed and less centralized. With the diffuse, broad set of problems with which we see active engagement, the value of a connective agent cannot be overstated. A bridge-building organization that connects people and communities to build a robust network can also achieve a significant impact by extending the coalition of partners invested these efforts.

With this vision in mind, we see both promise and inspiration in Cabaj & Weaver’s “Container for Change”:

Building a strong container requires paying attention to a variety of dimensions of backbone stewardship. Some of the more important ones are the following:

- The mobilization of a diverse group of funders, backbone sponsors, and stewardship arrangements that demonstrate cross-sectoral leadership on the issue.
- The facilitation of the participants’ inner journey of change, including the discovery and letting go of their own mental models and cultural/emotional biases, required for them to be open to fundamentally new ways of doing things.
- Processes to cultivate trust and empathy amongst participants so they can freely share perspectives, engage in fierce conversations, and navigate differences in power.
- Using the many dilemmas and paradoxes of community change — such as the need to achieve short-term wins while involved in the longer-term work of
system change – as creative tensions to drive people to seek new approaches to vexing challenges without overwhelming them.

- Timely nudges to sustain a process of self-refueling change that can sustain multiple cycles of learning and periodic drops in momentum and morale.\textsuperscript{34}

Adopting this approach does not diminish the opportunity for a more formally coordinating backbone organization to emerge in a subcommunity working within this space. Moreover, it does not eliminate the need for those interested in seeing an improved future state of affairs to collectively recognize and commit to reasonable funding for an organization to undertake and continue this work.

\textsuperscript{34} Cabaj & Weaver, 10-11
NEXT STEPS

WHAT’S NEXT FOR THE PEGI PROJECT?

During 2018, the PEGI Project convened stakeholders from libraries, academia, non-profit, scientific, and government sectors to surface a broadly shared national agenda on future steps needed to provide long-term access to born-digital government information. Among a diverse group of experts and invested organizations, we found a common sense of urgency, a collective commitment to grow and extend work already underway, and a shared understanding of the need to extend coalition-building efforts to include those whose investment can move forward broader and more coordinated efforts.

An extensive period of reflection and deliberation has led the PEGI Project team to identify the following future directions, which we think align well with strategic opportunities to continue to create change at a national level.

1. The PEGI Project as connective agent

Both the immediate and sustained interest in our work has demonstrated the need for an organization that can act as a vital and responsive connector among the many active efforts within this space. Time and again, we have found that fostering connections leads to an improved sense of community and commonality with those whose work is otherwise dispersed — geographically, technologically, and professionally. To mobilize cross-sectoral leadership within this diffuse community, we must continue to seek and build trust among potential partners.

Taking inspiration from the proposals created in the “Pathways Forward” activity at the National Forum, we plan to extend our efforts to connect those who want to participate in efforts, with those who are
organizing and accomplishing meaningful work. In particular, we will continue to focus on seeking engagement with projects that support underserved communities, and those led by practitioners from diverse and underrepresented backgrounds. In expanding the scope of our community engagement efforts, we also hope to more directly contribute to a more inclusive environment for learning, collaborating, and nurturing new projects and activities.

Our strength in building these relationships lies in our growing expertise about the problem-space for long-term access to born-digital government information, as well as our focus on actively listening to the interests and concerns of those with whom we work. All of us are committed to helping our colleagues strengthen and refine their own efforts, so they can become more effective and better leverage available opportunities.

2. The PEGI Project as educational partner

As government information librarians, we have found new insight into challenges that often demarcate professional boundaries between libraries and archives. We believe there is a significant and timely opportunity to develop and provide access to training that seeks to bridge communication and increase the shared understanding of essential work accomplished by those working in both types of institutions.

To leverage this opportunity, we plan to seek funding that will lead to the development of a curriculum and toolkit to build community engagement with electronic government information preservation, targeted for both new librarians and current professionals. We intend to seek partnerships with archivists, records managers, instructional designers, and library educators to create training and tools to support the development of records management knowledge among librarians, particularly those with interest or experience in government information, data, or e-government services, but who have less exposure to archives. Outputs from this work will also be designed to support the initiation of sustainable, locally-focused collection activities that complement existing efforts underway, including those in archives, libraries, and agencies. Access to this training will, first and foremost, increase effectiveness in supporting researchers and the general public. We hope it will also kickstart additional institutional efforts that lead to the expansion of existing communities of practice.

Building awareness related to records management, particularly as it pertains to government agencies, can also increase awareness of how government information and data are preserved for long-term access and use. In doing so, we also hope to increase awareness in government records management communities about the interplay between the government’s interest and the public’s interest in decisions made throughout the appraisal process. In the process of developing the proposed curriculum, we could, for example, convene targeted table conversations to seek common ground that leads to professional dialogue about the continuing evolution of records management practices.

3. The PEGI Project as leadership catalyst

We see a significant opportunity to encourage increased investment on the part of academic and research libraries. Many libraries are already developing capacity for curating research products on
behalf of their parent institutions. By developing a message that resonates with library administrators, we hope to activate and catalyze efforts to collect and curate born-digital government information and data that align with organizational missions and, at the same time, improve preservation and access for these vital resources.

To accomplish this, we hope to engage a communications expert to help us craft our message to better resonate with those who can initiate or reinforce efforts that complement those already underway. The many stakeholders with whom we have already connected have affirmed our own sense that this cause is compelling. We believe that coalition-building efforts will be strengthened among a broader group of stakeholders if we are able to communicate the urgency of this work to those who can direct existing efforts or undertake new projects aligning with these needs. Experts we engaged with agreed that grassroots efforts could become more effective by adopting a more standard approach to describing their activities. A lightweight communications toolkit could serve as common grounds for a starting point.

To continue to build on the weight of evidence we have at hand, we also plan to build on the research accomplished in the 2018 PEGI Environmental Scan by articulating the need for a research agenda. This would include reviews of ‘lost’ or fragile information, the universe of available information, and risks and vectors for loss. Products of the research agenda can more firmly highlight the disconnect between the future needs of academic research and current data management and curation practices.

Finally, we plan to extend our success in bridging mutual interests by initiating a series of targeted presentations and invited table conversations to engage library administrators with issues surfaced in our work to date. We intend to develop strategies to motivate active collection and preservation activities on the parts of their institutions, with the insight and input that emerge from these events.
As we have explored above, we see ways that we can learn from the Collective Impact model as the PEGI Project moves forward. Cabaj & Weaver (2016) offer a revised version of the model that aligns more realistically with the diffuse, amorphous communities of practitioners who are engaged with electronic government information. The work the PEGI Project has accomplished demonstrates a shared aspiration, if not a formal agenda, to achieve better preservation of electronic government information. We have made strides in creating learning opportunities and in meaningfully engaging and connecting with other groups working for better preservation and access.

Collectively, we also recognize that there are gaps to fill for our efforts going forward. Though our focus has tended toward the federal government information ecosystem, we need to also engage with local, state, and tribal government information stakeholders. We know information from these jurisdictions can be at significant risk of loss, and frequently receives less attention than federal information. These stakeholders also have insight into the practice and policy issues that affect the federal government. Similarly, there are gaps in our explorations of the user communities that we intend to address. In our series of mini-forums, the PEGI Project concentrated on connecting with academic researchers; however, there are many public interest groups and non-governmental organizations that use government information in their work. These organizations set priorities for community development and help to shape policy, much of which is based on data, analysis, and other government-produced information.

The gaps also extend to the PEGI Project steering committee. To date, the team is a close-knit group of librarians, most of whom serve or have served as Federal Depository Library coordinators. Both Kania & Kramer’s original Collective Impact model and Cabaj & Weaver’s revised version place emphasis on cross-sectoral leadership. Knowing that leadership is key in continuing to build momentum around preserving electronic government information, we recognize the need to bring into the steering committee members from other sectors who are actively working for change in this area. However, we also want to be strategic in expanding the group to provide the influence that can move the Project forward.

The IMLS National Leadership Grant provided opportunities that would have otherwise been out of reach for us. Thanks to its support, we organized productive meetings involving a variety of stakeholders. The knowledge gained from these activities helped us to refine our approach to using the Collective Impact model as the project framework. Although we expect more growth and change over the coming years, our core mission will remain the same. We will continue to educate and build aspirations for community engagement as we work toward better preservation and access to electronic government information.
PEGI PROJECT: Timeline of Project Activities

Activities marked with ** were included in the IMLS grant as activities supporting the Preservation of Electronic Government Information (PEGI) Project. Other activities further support and raise awareness for the project. Most of these activities were funded by the participant’s institution.

May 2016: Digital Preservation of Federal Information Summit, San Antonio, Texas

Held in conjunction with the Coalition of Networked Information (CNI) spring meeting and facilitated by Dr. Katherine Skinner of the Educoopia Institute, the Digital Preservation of Federal Information Summit brought together archivists, librarians, technologists, and others to engage in a structured dialogue on at-risk digital government information. It also aimed to explore the development of a national agenda to address the preservation of access to electronic government information.

Read the summit report: https://digital.library.unt.edu/ark:/67531/metadc826639/


Held in conjunction with the Coalition of Networked Information (CNI) fall meeting and facilitated by Michelle Gallinger of Gallinger Consulting, this meeting brought together a small group of librarians and archivists to discuss and scope a grant proposal to address concerns of at-risk electronic government information. At this meeting a core group committed to serve on the PEGI project team.

May 2017: Open Access Symposium, Frisco, TX

A moderated panel featuring Mark Phillips of the University of North Texas (UNT), David Walls of the Government Publishing Office (GPO), Margaret Janz from the University of Pennsylvania, and Shari Laster, then of the University of California, Santa Barbara (UCSB), explored current preservation activities and issues impacting work in this area.

View the archived files: https://openaccess.unt.edu/symposium/2017/preservation-and-access-government-information-and-data-accounts-field

May 2017: PEGI Field Work, Washington, DC

With funding from the University of North Texas (UNT), Roberta Sittel and Deborah Caldwell...
interviewed representatives within federal agencies to better understand publishing workflows, records scheduling and management with the National Archives and Records Administration (NARA), and coordination with the Government Publishing Office (GPO) for wider distribution. Interviews were conducted with the Marine Mammal Commission, the National Archives and Records Administration (NARA), Library of Congress, and the Government Publishing Office (GPO).

October 2017: Practice Mini-Forum at the Federal Depository Library Conference, Arlington, VA

In October 2017, the PEGI Project Team converged at GPO’s FDLP annual meeting in Arlington, Virginia to facilitate a mini forum for government information librarians. PEGI leaders pitched a series of questions designed to explore the nature and breadth of problems concerning preservation of born-digital government information. Participants were seated in small groups, and each table took an interesting and unique direction.

- Table notes from Marie Concannon: https://www.pegiproject.org/blog/2018/1/17/pegi-dlc-meeting-table-notes-marie-concannon
- Table notes from Scott Matheson: https://www.pegiproject.org/blog/2018/2/18/pegi-dlc-meeting-table-3-recap


PEGI Project team members participated in two presentations at the 2017 Digital Library Federation (DLF) Forum. Shari Laster, then at UCSB, Martin Halbert at University of North Carolina Greensboro, and Katherine Skinner of the Educopia Institute presented, “Mobilizing Preservation of Electronic Government Information in Fraught Times.” Shari Laster also participated in a panel discussion entitled “Government Records and Data Transparency in the Digital Age: Describing the Problem, Imagining Solutions,” facilitated by Rachel Mattson at University of Minnesota, with Emily Knox at University of Illinois Urbana-Champaign and Britt Paris of the Data & Society Research Institute.

Later that week, PEGI Project team members Martin Halbert, Katherine Skinner, and Lynda Kellam also gave a similar presentation at the adjacent Digital Preservation (DigiPres) conference.

January 2018: Mini-Forum at the American Historical Association (AHA) Annual Meeting, Washington, DC

The first of the IMLS-funded mini-forums was held at the annual meeting of the American Historical Association (AHA). This was the first discussion during which the team engaged participants with our set of three questions. Discussion participants included students, practicing historians, and government employees. The discussion focused on records related to government decision-making, such as emails and internal memos. Discussion also centered around the risk of materials from the mid-1990s to mid-2000s.
February 2018: Mini-Forum at the American Library Association (ALA) Mid-Winter Meeting, Denver, CO**

At ALA Midwinter, the PEGI Project, utilizing the three questions, facilitated an active discussion with forum attendees, including several dozen members of ALA’s Government Documents Round Table (GODORT). Much of this discussion reflected similar concerns and insights as the FDLC mini-forum.

February 2018: Mini-Forum at the American Association for the Advancement of Science (AAAS) Annual Meeting, Austin, TX**

In an attempt to engage with researchers and practitioners from the science community, the PEGI Project organized a mini-forum at the AAAS annual meeting in Austin, Texas. The mini forum drew no participation, though PEGI team members were able to engage AAAS attendees between and during sessions. A lack of familiarity with this professional group, coupled with a lack of advertising for the session could account for the lack of participation. This session served as a lesson in the team’s planning and execution of future mini-forums, including best use of grant dollars to expected participation.

March 2018: PEGI Project Virtual Mini-Forum**

This virtual mini-forum allowed interested participants to consider and reply to the three questions. Attendees included librarians, archivists, students, and government employees, with a wide range of disciplinary expertise. Important opportunities for coordinated effort emerged during group discussion and brought a number of projects to our attention as possible models for cooperative activity.

View the recording of the webinar: https://www.pegiproject.org/webinars/

March 2018: Poster Presentation at the Research Data Access & Preservation (RDAP) Summit, Chicago, IL

Team member Marie Concannon of the University of Missouri presented a poster promoting the PEGI Project at the Research Data Access & Preservation Summit. The poster highlights the work of the project and early findings from the team’s discussions with stakeholder communities.

View the poster: https://osf.io/vzx9s/

April 2018: PEGI at the Organization of American Historians (OAH) Annual Conference, Sacramento, CA**

The PEGI Project’s participation at the OAH Annual Conference focused on tabling in the exhibit hall to raise awareness of the need for a coordinated strategy for collecting and preserving born-digital government information. We spoke with independent scholars, vendors, and other attendees about the current landscape for this work, and came away with more insight on researcher perspectives.

Due to budgetary reasons and timing of activities, this event substituted for participation in the Research Data Alliance (RDA) and International Data Week (IDW) co-conference in Gaborone,
Botswana.

**May 2018: PEGI Project Webinar**

For the second virtual mini-forum, we invited key activists and stakeholders to reflect on the PEGI Project “three questions.” Presenters included:

- Rachel Mattson, Curator of the Tretter Collection for GLBT Studies at the University of Minnesota Libraries & founder/co-leader of the Digital Library Federation’s interest group on Government Records Transparency & Accountability
- Brandon Locke, then Director of LEADR at Michigan State University & founder & co-organizer of Endangered Data Week
- Bernard F. Reilly, then President of the Center for Research Libraries
- Justin Schell, Director of Shapiro Design Lab & member of the Environmental Data & Governance Initiative (EDGI)
- Bethany Wiggin, Founding Director of the Penn Program in Environmental Humanities
- Patricia Kim, then Program Coordinator of the Penn Program in Environmental Humanities
- Heather Joseph, Executive Director, SPARC

These speakers discussed highly relevant approaches and considerations for motivating and organizing efforts to preserve government information. The long-term cost of inaction is certainly one motivator, but both the scale of the challenge and the invisible nature of the infrastructure needed to address it can make it hard to know how to proceed with any given solution. Seeking out collective investment in data at a more personal, human-navigable scale can help to push through some inertia and create avenues of engagement. Building coalitions with projects leading change in related areas of concern, such as Open Access, can amplify progress that makes our own efforts more likely to succeed.

View the recording of the discussion: [https://www.pegiproject.org/webinars/](https://www.pegiproject.org/webinars/)

**May 2018: Presentation at International Association of Social Science Information Services & Technology (IASSIST) Annual Conference, Montreal, Canada**

Project team member Lynda Kellam, then at the University of North Carolina-Greensboro, and Roberta Sittel (UNT) presented on early findings of the PEGI Project and encouraged IASSIST conference attendees to complete the three-questions survey.

**June 2018: Mini-Forum at the American Library Association (ALA) Annual Conference, New Orleans, LA**

Because the ALA Midwinter Meeting mini-forum in February 2018 was co-hosted by GODORT and focused on government information librarians and professionals, our mini-forum at the Annual Conference was targeted to reach conference attendees from other communities of practice, including research data management and scholarly communications. This meeting, which largely echoed prior discussions with the government information librarianship community as summarized
above, drew librarians from outside the United States, representatives of national libraries, and other professionals who do not necessarily specialize in government information but who recognize the importance of this issue in their own areas.

July 2018: Poster and Mini-Forum at the American Association of Law Libraries (AALL), Baltimore, MD

Scott Matheson, (Yale), and Roberta Sittel (UNT) presented a poster similar in focus to the one presented at RDAP earlier in the year.

We also conducted a mini-forum to raise awareness and gather input from the law library community. The well-attended session led to a lively discussion of past efforts and known access issues, and allowed us to gather additional perspective about legal information and the needs of law librarians, lawyers, and judges. Also connected with Legal Information Preservation Alliance (LIPA).

July 2018: Panel Presentation at the Association of Public Data Users (APDU) Annual Conference, Arlington, VA

Project team members Deborah Caldwell (UNCG) James R. Jacobs (Stanford), and Shari Laster from Arizona State University (ASU), collaborated with Margaret Janz (Penn) to give a presentation focusing on grassroots efforts to preserve born-digital government information, to an audience of government data stakeholders. Discussions at this conference helped to surface insights into policy considerations that affect data production, dissemination, and preservation.

August 2018: Mini-Forum at the Society of American Archivists (SAA), Council of State Archivists (SAA), and National Association of Government Archives & Records Administrators (NAGARA) Joint Meeting, Washington, DC**

At the joint meeting of the Society of American Archivists (SAA), the Council of State Archivists (CoSA) and the National Association of Government Archives and Records Administrators (NAGARA), the mini-forum featured lively, small-group discussions with archivists and records managers, many of whom had extensive experience with records management for state and local governments. Issues raised in this discussion had a significant impact on how we framed discussions engaging with librarians and archivists going forward.

September 2018: Panel Presentation at the International Conference on Digital Preservation (iPRES), Boston, MA**

Project participation at the International Conference on Digital Preservation (iPRES) included a panel discussion focusing on how policies affect access to and preservation of information. The panel, which included David Walls and Cindy Etkin from GPO, Carla Graebner from Simon Fraser University, Frederick Zarndt, a preservation consultant, and Edward McCain with the Reynolds Journalism Institute, also discussed collaborative approaches to addressing policies that are both positive and negative for the preservation of electronic government information. Speakers from GPO explained how policies guide programs and the importance of engagement with policy-setting. Speakers with international experience highlighted policies that have helped and hindered electronic information preservation efforts. This session reinforced the importance of raising awareness about the need for participation in setting good policy.

Scott Matheson (Yale) presented a paper on issues of concern to the PEGI Project at the APS annual conference, inspired by the conference theme of the Past, Present and Future of Libraries. This paper is now forthcoming in the 2018 conference proceedings.

September 2018: “Planning a Community-Created Data Rescue Toolkit” Planning Meeting, Baltimore, MD

Deborah Caldwell (UNCG) and Shari Laster (ASU) accepted an invitation to attend a stakeholder meeting for the IMLS funded planning grant “Planning a Community-Created Data Rescue Toolkit,” hosted at Johns Hopkins University. In two days of discussion, we contributed to consideration and planning of a potential toolkit, and built important connections with experts who we later invited to attend the PEGI Project National Forum.

November 2018: Panel Presentation at Beyond the Numbers, St. Louis, MO

Marie Concannon, from the University of Missouri, and Lynda Kellam (UNCG) partnered with Katrina Stierholz of the Federal Reserve Bank of St. Louis to discuss preservation of born-digital economics information, including data sources.

December 2018: PEGI Project National Forum, Washington, DC

The PEGI National Forum, held December 9-10, 2018 in Washington DC, was the capstone to our series of mini-forums. Whereas mini-forums were planned as 1-2 hour sessions for the purpose of gathering responses to our three questions as posed to specific communities, the National Forum was a two-day event designed to bring together diverse groups to take a deeper dive into the issues. National Forum participants were selected based on their track record of leadership, experience, or interest in electronic government information preservation, as well as for their ability to effectively represent their respective fields and industries. This forum brought together key stakeholders from federal agencies, state government leaders, open government advocates, archivists, librarians, policy analysts, data specialists, scientists, journalists, digital preservationists, historians, and others.

For a detailed summary of this event, see the National Forum Report: https://educopia.org/pegi-project-national-forum-summary-and-report/

March 2019: “The PEGI Project Takes Flight” webinar for Help! I’m an Accidental Government Information Librarian

The project team provided the Help! series and library community an update on the project. The webinar offered a recap of the National Forum and initial impressions of the Collective Impact model as explored by the PEGI Project.

View the recording of this webinar: https://www.youtube.com/watch?v=JURSoOCoDrk

May 2019: PEGI Project Webinar

In May 2019, the project team presented a webinar to share key outcomes from the National Forum, and discuss potential directions for future efforts.
View the recording of this discussion: https://youtu.be/yefELGFB1ac

June 2019: Presentation at American Library Association Annual Conference, Washington, DC

In a program sponsored by the ALA Government Documents Round Table (GODORT), Martin Halbert (UNCG), Roberta Sittel (UNT), Shari Laster (ASU), James R. Jacobs (Stanford), and Deborah Caldwell (UNCG) discussed initial project outcomes and reflected on future directions as influenced by the Collective Impact model.


Scott Matheson (Yale) and Deborah Caldwell (UNCG) presented a poster sharing reflections on the Collective Impact model as it pertains to project efforts. Law librarians and government librarians continued strong interest in and concern for the preservation of electronic information.

August 2019: Poster at the Society of American Archivists Annual Conference, Austin, TX

Roberta Sittel (UNT) and Deborah Caldwell (UNCG) presented a similar poster reflecting on the Collective Impact model, and engaged in discussions with attendees. Government records managers shared the need for an agency-led, top-down approach for records management and preservation. Attendees talked of records that are of interest to the agency and potential researchers outside of NARA’s scope; citing a need for agency directed preservation of these materials. These conversations indicate additional need for the PEGI Project team to pursue further discussion with archivists and records managers at multiple levels of government agencies.

Forthcoming in 2019

- Shari Laster (ASU), James R. Jacobs (Stanford), and Deborah Caldwell (UNCG) have been accepted to present a short talk on our efforts to engage with stakeholder communities at the Digital Preservation 2019 conference, hosted by the National Digital Stewardship Alliance in October 2019 in Tampa, FL.
- The project team will share an update with the Federal Depository Library Program community at GPO’s Federal Depository Library Conference in October 2019, in Arlington, VA.


https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/OMB/circulars/a130/a130revised.pdf [https://perma.cc/72XN-6CM4].

https://digital.library.unt.edu/ark:/67531/metadc85399/ [https://perma.cc/U36Y-8HUH].


